

CC GUS may be used to achieve extracellular detoxification of glucuronides
 CC (e.g. toxin glucuronide) and to examine conjugation patterns of
 CC glucuronides. Microbial GUS may also be used in traditional medical
 CC diagnostic assays, for drug testing, pharmacokinetic studies,
 CC bioavailability studies, diagnosis of diseases and syndromes, following
 CC progression of disease or its response to therapy. Microbial GUS has
 CC increased thermal stability, high turnover number and enzymatic activity.
 CC It is highly specific for the substrate and water soluble, and the
 CC substrates are stable

XX Sequence 563 AA;

Query Match 99.9%; Score 2999; DB 3; Length 563;
 Best Local Similarity 100.0%; Pred. No. 1e-236;
 Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRPQNNKKRPFILINGVNLVETSKDRPIAVPGSMNQYDLCYEEGPFYKTFYVVK 60
 DB 1 MVRPQNNKKRPFILINGVNLVETSKDRPIAVPGSMNQYDLCYEEGPFYKTFYVVK 60
 QY 61 XLSQKHRLVFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTVGKVGSENELRVVENRL 120
 DB 61 XLSQKHRLVFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTVGKVGSENELRVVENRL 120
 QY 121 KVGSPSKVPDPSGTHTVGFFSGPPANDFPPYGGIIRPVLIETPTDARILIDIVDTSES 180
 DB 121 KVGSPSKVPDPSGTHTVGFFSGPPANDFPPYGGIIRPVLIETPTDARILIDIVDTSES 180
 QY 181 EPEKLGKVKYKIEVSEBAVQEMTKLGESEKKIRTSNRFVEGEFTLENARFWSLEDPY 240
 DB 181 EPEKLGKVKYKIEVSEBAVQEMTKLGESEKKIRTSNRFVEGEFTLENARFWSLEDPY 240
 QY 241 LYPKVELEKDEYTLDGIRITISWDEKRLVNGKRPVLKGGKHEEPVVGQGTFFYLM 300
 DB 241 LYPKVELEKDEYTLDGIRITISWDEKRLVNGKRPVLKGGKHEEPVVGQGTFFYLM 300
 QY 301 KDFNLKMINANSFRTSHYPSSEBMLDLADRLGILVIDEAPHVIGITRYHNPETOKIAED 360
 DB 301 KDFNLKMINANSFRTSHYPSSEBMLDLADRLGILVIDEAPHVIGITRYHNPETOKIAED 360
 QY 361 NIRMIRHKNHPSVIMSVANEPESNHPDAEGFKALYETANEMDTRPVVMSMDAP 420
 DB 361 NIRMIRHKNHPSVIMSVANEPESNHPDAEGFKALYETANEMDTRPVVMSMDAP 420
 QY 421 DERTDVALKXFYDVCNRYGYWYIYOGRIEESGLALEKDIIBELYARHRKPIFVTEGAD 480
 DB 421 DERTDVALKXFYDVCNRYGYWYIYOGRIEESGLALEKDIIBELYARHRKPIFVTEGAD 480
 QY 481 AIAGIHYDPQMFSEBYOAEIYEKTIIRLLKKOYIIGTHVWAPADFTTPONVRRPILNKH 540
 DB 481 AIAGIHYDPQMFSEBYOAEIYEKTIIRLLKKOYIIGTHVWAPADFTTPONVRRPILNKH 540
 QY 541 GVFTDRQPKLVAVHLRLMSEV 563
 DB 541 GVFTDRQPKLVAVHLRLMSEV 563

RESULT 2
 AAM93820
 ID AAM93820 standard; protein; 602 AA.
 XX
 AC AAM93820;
 XX
 DT 25-JUN-1999 (first entry)
 XX
 DE Bacillus sp. GUS protein.
 XX
 KW GUS; beta-glucuronidase; secreted; reporter molecule; marker;
 KW receptor molecule; diagnostic tool; transgene construction; plant;
 XX insect; cleavage; detoxification; glucuronide.
 OS
 XX Bacillus sp.

PN W09913085-A2.
 XX 18-MAR-1999.
 PD
 XX 09-SEP-1998; 98WO-US019217.
 PF
 XX 09-SEP-1997; 97US-0058263P.
 PR
 XX (CAMP-) CAMBIA BIOSYSTEMS LLC.
 PA
 XX Jefferson RA, Kilian A, Keese PK;
 PI
 XX WPI; 1999-229241/19.
 DR
 XX New isolated microbial beta-glucuronidase.
 PT
 XX Claim 4; Fig 3; 76pp; English.
 PS
 XX

CC This invention describes a novel secreted form of Bacillus sp. beta-
 CC glucuronidase (Bogus). The microbial Bogus polypeptide can be used as a
 CC reporter/effector molecule and as a diagnostic tool. The products of the
 CC invention can be used as markers for transgene constructions, e.g. in
 CC plants or insects. They can also be used for the cleavage and
 CC detoxification of glucuronides and to examine conjugation patterns of
 CC glucuronides

XX Sequence 602 AA;

Query Match 31.1%; Score 932; DB 2; Length 602;
 Best Local Similarity 35.6%; Pred. No. 2.7e-67;
 Matches 221; Conservative 94; Mismatches 223; Indels 82; Gaps 15;

QY 1 MVRPQNNKKRPFILINGVNLVETSKDRPIAVPGSMNQYDLCYEEGPFYKTFYVVK 44
 DB 1 MVRPQNNKKRPFILINGVNLVETSKDRPIAVPGSMNQYDLCYEEGPFYKTFYVVK 44
 QY 45 YEBGPFYKTFYVVKXLSQKHRLVFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTVGK 104
 DB 45 YEBGPFYKTFYVVKXLSQKHRLVFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTVGK 104
 QY 61 NHIQYVYREFTVPAVYLKQRIIVRPSATHTKALVYVNGSLVHEKGGFLPEEAINNS 120
 DB 61 NHIQYVYREFTVPAVYLKQRIIVRPSATHTKALVYVNGSLVHEKGGFLPEEAINNS 120
 QY 105 VSGSENELRVVENRLKVGSPSKVPDPSGTHTVGFFGS-----FPANPDPFFPYG 154
 DB 105 VSGSENELRVVENRLKVGSPSKVPDPSGTHTVGFFGS-----FPANPDPFFPYG 154
 QY 121 LRDGMKRVYVAVDNIL-----DSTLPVGLYSEHREGKVIIRNKNPDPFENYA 170
 DB 121 LRDGMKRVYVAVDNIL-----DSTLPVGLYSEHREGKVIIRNKNPDPFENYA 170
 QY 155 GIIRPVLIETPTDARILIDIVDTSESSEPKLGGKVKVIEVSEBAVQEMTKLGESE 212
 DB 155 GIIRPVLIETPTDARILIDIVDTSESSEPKLGGKVKVIEVSEBAVQEMTKLGESE 212
 QY 171 GLHRPVKIYTFPTFYVEDISVTDENG---TGVTYTVDFQ---GAEYVKAIVVDEE 223
 DB 171 GLHRPVKIYTFPTFYVEDISVTDENG---TGVTYTVDFQ---GAEYVKAIVVDEE 223
 QY 213 KIRTSNRFVEGEFTLENARFWSLEDPYLPKVELEKDEYTLDGIRITISWDEKRLV 267
 DB 213 KIRTSNRFVEGEFTLENARFWSLEDPYLPKVELEKDEYTLDGIRITISWDEKRLV 267
 QY 224 GKVASTBESLGSNVAIIPVILMEPLNTLYLQIKVNLVDGLTIDYEEBPGVRYEVNDG 283
 DB 224 GKVASTBESLGSNVAIIPVILMEPLNTLYLQIKVNLVDGLTIDYEEBPGVRYEVNDG 283
 QY 266 RLYLNGKRPVLKGGKHEEPVVLGQGTFFYPLMIKDFNLKMINANSFRTSHYPSSEWLD 327
 DB 266 RLYLNGKRPVLKGGKHEEPVVLGQGTFFYPLMIKDFNLKMINANSFRTSHYPSSEWLD 327
 QY 284 KFLINNKPFYFGKGEHEDTPINGRGENASVMDNFIKMGANSFRTAHYPSSELMR 343
 DB 284 KFLINNKPFYFGKGEHEDTPINGRGENASVMDNFIKMGANSFRTAHYPSSELMR 343
 QY 328 LADRLGILVIDEAPHVIGITRYHYN-----PEQOKIA-----ENIRMRIDR 368
 DB 328 LADRLGILVIDEAPHVIGITRYHYN-----PEQOKIA-----ENIRMRIDR 368
 QY 344 LADREGVLVIDETPAVG---HLNFMATTLGEGSERVSTWEKTRTEHHODVNLSELVS 400
 DB 344 LADREGVLVIDETPAVG---HLNFMATTLGEGSERVSTWEKTRTEHHODVNLSELVS 400
 QY 369 HKNRPSVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVVMVSMMDAPDERTDV 427
 DB 369 HKNRPSVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVVMVSMMDAPDERTDV 427
 QY 401 DKNRPSVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVVMVSMMDAPDERTDV 458
 DB 401 DKNRPSVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVVMVSMMDAPDERTDV 458
 QY 428 ALKTFDYCVNRYGYWYIYOGRIEESGLALEKDIIBELYARHRKPIFVTEGADAIA 483
 DB 428 ALKTFDYCVNRYGYWYIYOGRIEESGLALEKDIIBELYARHRKPIFVTEGADAIA 483
 QY 459 VAEIYDVALNRKYGWYFDGDL---AAKVRHLOEFNAMKRCGKPKIMITEYGADIVA 515
 DB 459 VAEIYDVALNRKYGWYFDGDL---AAKVRHLOEFNAMKRCGKPKIMITEYGADIVA 515
 QY 484 GIHYDPQMFSEBYOAEIYEKTIIRLLKKOYIIGTHVWAPADFTTPONVRRPILNKH 543
 DB 484 GIHYDPQMFSEBYOAEIYEKTIIRLLKKOYIIGTHVWAPADFTTPONVRRPILNKH 543
 QY 516 GFHDIDPMTFEEYQVYQANHVVFDEFENFVGBQANMFADFAISQGVWVQGNKGVF 575
 DB 516 GFHDIDPMTFEEYQVYQANHVVFDEFENFVGBQANMFADFAISQGVWVQGNKGVF 575
 QY 544 TRDRQPKLVAVHLRLMSEV 563
 DB 544 TRDRQPKLVAVHLRLMSEV 563

```
DB      |||:|||||:|:|
576 TRDRKPKLAHVFRERWTNI 595

RESULT 3
AAW93825
ID      AAW93825 standard; protein; 602 AA.
XX
AC      AAW93825;
XX
DT      25-JUN-1999 (first entry)
XX
DE      Bacillus sp. codon optimised GUS protein.
XX
KW      GUS; BcGUS; beta-glucuronidase; secreted; reporter molecule; marker;
KW      receptor molecule; diagnostic tool; transgene construction; plant;
KW      insect; cleavage; detoxification; glucuronide.
XX
OS      Bacillus sp.
XX
PN      MO9913085-A2.
XX
PD      18-MAR-1999.
XX
PF      09-SEP-1998; 98WO-US019217.
XX
PR      09-SEP-1997; 97US-0058263P.
XX
PA      (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI      Jefferson RA, Killian A, Keese PK.
XX
DR      WPI; 1999-229241/19.
XX
DR      N-PSDB; AAX23826.
XX
PT      New isolated microbial beta-glucuronidase.
XX
PS      Example 3; Fig 13A-D; 76pp; English.
XX
CC      This invention describes a novel secreted form of Bacillus sp. beta-
CC      glucuronidase (BcGUS). The microbial BcGUS polypeptide can be used as a
CC      reporter/effector molecule and as a diagnostic tool. The products of the
CC      invention can be used as markers for transgene constructions, e.g. in
CC      plants or insects. They can also be used for the cleavage and
CC      detoxification of glucuronides and to examine conjugation patterns of
CC      glucuronides
CC
SQ      Sequence 602 AA;

Query Match      31.0%; Score 929; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 4,8e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

OY      1 MVRPQRNKKRFLILNGVWNLV-----TSKORPIAVPGSMNE--QYDLC 44
DB      1 MLYPNTETRGVFDLNGVWNLFDYKGLSEKWEYSKLTDTISMAVPSYNDIGYTKIR 60
OY      45 YEEGFYKTFYVYPKLSOKIIRLYFAAVNTDCEVPLNGEKVGNHIEYLPFEVDVYTK 104
DB      61 NHIGVWYEREFYVAYLKQRIVLRFSGATHKAIYVYNGELVYHKGFLPFEELINNS 120
OY      105 VKSGENELRVVVENRLKVGSPSKVPDSGTYTVGFFSG-----PPANDFPPYG 154
DB      121 LRDGNRVTVAVNDIL-----DSTLPLVGLYSRHEGLGKVI RNKNPNDFPFMYA 170
OY      155 GIIRPVLEFTHARILDIWYDTSESEBEKLGKVKVIVSEAVGEMTIKG--EE 212
DB      171 GLHRPVKITYTPFYVEDISVVDPNGP--TGVTYTVVDPO---GAAETVKQSVVDEE 223
OY      213 KIRTSNRFVEGEFILLENARFWSLEDPYLYPLKVLKELKDEYTLDI-----GIRTSMDKE 267
DB      224 GKVAVSTGSLGSGNVAEIPVILMEPLNTLYLQIKVELVNDGLTIDYVEEPFGVRYEVNDG 283
```

```
OY      268 RLYNGKRVFLKRGKEHEEPVVGCTFYPLMIXDNLLKMINANSFRSSHYPYSEEMLD 327
DB      284 KFLINNKPFYFGKGEKHEDTPINGGFNEASNVMDNLIKIGANSFRTAHYPYSEELMR 343
OY      328 LADRGLIVIDEAPHVIGITRYHYN-----PETOKA-----EDNIRRMIDR 368
DB      344 LADREGLVVIDETRAVGV---HLNFMATYTGDEGSESRVSTWEKIRTFEHHQDVLRELYSR 400
OY      369 HKNHPSVIMWSVANEPESNHNPDAEGFFKALYETANEMD-RTRPVVWSMDAPDERTDV 427
DB      401 DKNHPSVVMWSIANEPAATEEGAYEFKPLVELTKELDPOKRPTVIVLFVMAATPE--TDK 458
OY      428 ALKTYDLYCVNRYYGMYTYQGRIEGLQALEKDIEELYARRH----KPIFTEPADALIA 483
DB      459 VAEILDVIALNRKMGWYFDGDLLEAKVHLR---OEFAMNKKRCGKPDIMITEYGADTV 515
OY      484 GIHYDPQMFSEBEYOAEIYEKTIIRLLKKDYIITGHVAFADFCKTPQVVRPILNHKGVF 543
DB      516 GFHDIDPVMFTBEYQVEYTYQANHYVDFEFENFVGEQAMNFADFSQGVMYQGNKGVF 575
OY      544 TRDRQPKLVAHVLRRLMSEV 563
DB      576 TRDRKPKLAHVFRERWTNI 595
```

```
RESULT 4
AAW93822
ID      AAW93822 standard; protein; 602 AA.
XX
AC      AAW93822;
XX
DT      25-JUN-1999 (first entry)
XX
DE      Bacillus sp. GUS protein.
XX
KW      GUS; BcGUS; beta-glucuronidase; secreted; reporter molecule; marker;
KW      receptor molecule; diagnostic tool; transgene construction; plant;
KW      insect; cleavage; detoxification; glucuronide.
XX
OS      Bacillus sp.
XX
PN      MO9913085-A2.
XX
PD      18-MAR-1999.
XX
PF      09-SEP-1998; 98WO-US019217.
XX
PR      09-SEP-1997; 97US-0058263P.
XX
PA      (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI      Jefferson RA, Killian A, Keese PK;
XX
DR      WPI; 1999-229241/19.
XX
PT      New isolated microbial beta-glucuronidase.
XX
PS      Example 1; Fig 5; 76pp; English.
XX
CC      This invention describes a novel secreted form of Bacillus sp. beta-
CC      glucuronidase (BcGUS). The microbial BcGUS polypeptide can be used as a
CC      reporter/effector molecule and as a diagnostic tool. The products of the
CC      invention can be used as markers for transgene constructions, e.g. in
CC      plants or insects. They can also be used for the cleavage and
CC      detoxification of glucuronides and to examine conjugation patterns of
CC      glucuronides
CC
SQ      Sequence 602 AA;

Query Match      31.0%; Score 929; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 4,8e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
```

```

QY 1 MVRPQNRKKRFFILNGVWNLV-----TSKDRPIAVPGSNM--QYODLC 44
DB 1 MLYPINTETRGVFDLNGVWNLVFKLDYKGKLEKYESKLTDTISMAVSSINDIVGKIR 60
QY 45 YEEGPFTYKTFYVPKXLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTGK 104
DB 61 NHIGYVWYEREFVPAVLKQRIYLFQSGATHKAIYVNGELVHEHKGFLPFEALINNS 120
QY 105 VKSGENELRVVNRLLKVGFPSPKVPDSGTHTVGPGS-----PPANDFPPYG 154
DB 121 LRDMNRVTAAVNDIL-----DDSTLPVGLYSEHHEGLGKVIIRNKNPNDFFNYA 170
QY 155 GIIRPVLIETDARLIDIVDTSESEPEKLGKVKYKIESEBAGQEMTIKLG--EE 212
DB 171 GHRPVKIYTPPTFYVEDISVTPNGP--TGIVTYVDQ---GKAETVKSVDDE 223
QY 213 KKIIRTSNRFEVEGFIENARFMSLEDPYLPKYLEKDEYTLDI-----GIRTI 267
DB 224 GKVAATGEGLSGVEIPVILMEPLNTLYQIKVELVNDGLTIDIVEBPGVATVEYNDG 283
QY 268 RLYNGKRVFLKGGKHEEPVLQGTFFPLMIKDFNLKMIANSFRTSHYPSSEWLD 327
DB 284 KFLINNKRFYFGKFGHEDTPINGRGFNBSNVDFNLKMIANSFRTAHYPSSELMR 343
QY 328 LADRLGILVIDEAPHVIGITRYHN-----PETOKIA-----EDNIRRMIDR 368
DB 344 LADREGLVVIDETPAVG--HINFMATTTGEGSERVSTWEKIRTEHHQDVIRELVSR 400
QY 369 HKHNPVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVWVSMMDADPERTRDV 427
DB 401 DKHNPVWMSIANEATEEGAYEYFKPLVELTKELDPQRPVTIYLPVMAATPE--TDK 458
QY 428 ALKRPDIVCVNRYGYIYQRIEGLQALEKDIIELYARHR---KPIFTEFGADALA 483
DB 459 VAEILIVIALNRNGVYFDGDLDAKVHLR--QEFHAMNKRCPGPKPIMITEYGADTVA 515
QY 484 GIHDPDPQMFSEYQAEIVETIRLLKKDYIIGTHWAFADPKTPONVRRPILNHKGVF 543
DB 516 GFHIDIDVMTTEEYQVEYQANHVPDEFENFVGEQANMPADPATISQGVNRVQGNKGVF 575
QY 544 TRDRQPKLVAVHLRLMSEV 563
DB 576 TRDRKPKLAHVFRERWTNI 595

RESULT 5
AAB93826
ID AAB93826 standard; protein; 602 AA.
AC AAB93826;
XX
DT 25-JUN-1999 (first entry)
XX
DE Bacillus sp. GUS protein fragment.
XX
KM GUS; beta-glucuronidase; secreted; reporter molecule; marker;
KM receptor molecule; diagnostic tool; transgene construction; plant;
KM insect; cleavage; detoxification; glucuronide.
XX
OS Bacillus sp.
XX
PN WO913085-A2.
XX
PD 18-MAR-1999.
XX
PF 09-SEP-1998; 98MO-USO19217.
XX
PR 09-SEP-1997; 97US-0058263P.
XX
PA (CAMP-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Kilian A, Keesee PK;
XX

```

```

DR WPI; 1999-229241/19.
XX
PT New isolated microbial beta-glucuronidase.
XX
XX Example 3; Fig 13d; 76pp; English.
XX
CC This invention describes a novel secreted form of Bacillus sp. beta-
CC glucuronidase (BOGUS). The microbial BOGUS polypeptide can be used as a
CC reporter/effector molecule and as a diagnostic tool. The products of the
CC invention can be used as markers for transgene constructions, e.g. in
CC plants or insects. They can also be used for the cleavage and
CC detoxification of glucuronides and to examine conjugation patterns of
CC glucuronides
XX
SQ Sequence 602 AA;
QY 1 MVRPQNRKKRFFILNGVWNLV-----TSKDRPIAVPGSNM--QYODLC 44
DB 1 MLYPINTETRGVFDLNGVWNLVFKLDYKGKLEKYESKLTDTISMAVSSINDIVGKIR 60
QY 45 YEEGPFTYKTFYVPKXLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTGK 104
DB 61 NHIGYVWYEREFVPAVLKQRIYLFQSGATHKAIYVNGELVHEHKGFLPFEALINNS 120
QY 105 VKSGENELRVVNRLLKVGFPSPKVPDSGTHTVGPGS-----PPANDFPPYG 154
DB 121 LRDMNRVTAAVNDIL-----DDSTLPVGLYSEHHEGLGKVIIRNKNPNDFFNYA 170
QY 155 GIIRPVLIETDARLIDIVDTSESEPEKLGKVKYKIESEBAGQEMTIKLG--EE 212
DB 171 GHRPVKIYTPPTFYVEDISVTPNGP--TGIVTYVDQ---GKAETVKSVDDE 223
QY 213 KKIIRTSNRFEVEGFIENARFMSLEDPYLPKYLEKDEYTLDI-----GIRTI 267
DB 224 GKVAATGEGLSGVEIPVILMEPLNTLYQIKVELVNDGLTIDIVEBPGVATVEYNDG 283
QY 268 RLYNGKRVFLKGGKHEEPVLQGTFFPLMIKDFNLKMIANSFRTSHYPSSEWLD 327
DB 284 KFLINNKRFYFGKFGHEDTPINGRGFNBSNVDFNLKMIANSFRTAHYPSSELMR 343
QY 328 LADRLGILVIDEAPHVIGITRYHN-----PETOKIA-----EDNIRRMIDR 368
DB 344 LADREGLVVIDETPAVG--HINFMATTTGEGSERVSTWEKIRTEHHQDVIRELVSR 400
QY 369 HKHNPVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVWVSMMDADPERTRDV 427
DB 401 DKHNPVWMSIANEATEEGAYEYFKPLVELTKELDPQRPVTIYLPVMAATPE--TDK 458
QY 428 ALKRPDIVCVNRYGYIYQRIEGLQALEKDIIELYARHR---KPIFTEFGADALA 483
DB 459 VAEILIVIALNRNGVYFDGDLDAKVHLR--QEFHAMNKRCPGPKPIMITEYGADTVA 515
QY 484 GIHDPDPQMFSEYQAEIVETIRLLKKDYIIGTHWAFADPKTPONVRRPILNHKGVF 543
DB 516 GFHIDIDVMTTEEYQVEYQANHVPDEFENFVGEQANMPADPATISQGVNRVQGNKGVF 575
QY 544 TRDRQPKLVAVHLRLMSEV 563
DB 576 TRDRKPKLAHVFRERWTNI 595

RESULT 6
AAB28402
ID AAB28402 standard; protein; 602 AA.
AC AAB28402;
XX
DT 26-JAN-2001 (first entry)
XX

```


DE Staphylococcus beta-glucuronidase.
 XX Microbial, beta-glucuronidase; GUS; Enterobacter; Salmonella;
 KM Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
 KM transgenic insect; marker; glucuronide detoxification.
 XX
 OS Staphylococcus sp.
 XX WO200055333-A1.
 XX
 XX 21-SEP-2000.
 PD
 PF 16-MAR-2000; 2000WO-US007107.
 XX
 XX 17-MAR-1999; 99US-00270957.
 PR
 XX (CAMB-) CAMBIA BIOSYSTEMS LLC.
 PA
 XX Jefferson RA, Mayer JE;
 PI WPI; 2000-647075/62.
 DR N-PSDB; AAA07930.
 XX
 PT Novel microbial beta-glucuronidase genes and gene products used as
 PT reporter/effector molecule, as diagnostic tool, in positive selection, to
 XX target molecules to specific cells and to detect and track linked genes.
 PS Example 3; Fig 3; 116pp; English.
 XX
 XX The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
 CC genes were obtained from six different genera: Enterobacter/Salmonella,
 CC Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can
 CC be used as a reporter/effector molecule for transgenic constructions and
 CC in in vitro diagnostic applications. It may also be used to generate
 CC sentinel plants that serve as bioindicators of environmental status. It
 CC may be used to generate transgenic insects for tracking insect
 CC populations or to facilitate the development of a bioassay for compounds
 CC that affect molecules critical for insect development (e.g. juvenile
 CC hormone). Secreted GUS may also serve as a marker for beneficial fungi
 CC destined for release into the environment. In animal systems, secreted
 CC GUS may be used to achieve extracellular detoxification of glucuronides
 CC (e.g. toxin glucuronide) and to examine conjugation patterns of
 CC glucuronides. Microbial GUS may also be used in traditional medical
 CC diagnostic assays, for drug testing, pharmacokinetic studies, following
 CC bioavailability studies, diagnosis of diseases and syndromes, following
 CC progression of disease or its response to therapy. Microbial GUS has
 CC increased thermal stability, high turnover number and enzymatic activity.
 CC It is highly specific for the substrate and water soluble, and the
 CC substrates are stable
 XX
 SQ Sequence 602 AA;
 Query Match 31.0%; Score 929; DB 3; Length 602;
 Best Local Similarity 35.6%; Pred. No. 4,8e-67;
 Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

DB 224 GKVAATSEGLSGNVEIPVILMEPLNTLYLYIKVELVNDGLTIDYEEPFVRYTEVNDG 283
 QY 268 RLYVNGKRVPLKFGKGEKEEPVLAQGTFYPLMKIDPNLLKINANSFRTSHYPYSEEMLD 327
 DB 284 KFLINRKPFYKFGKEDPTINGRGEASVNDPFLIKKIGANSFRTAHYPYSEELMR 343
 QY 328 LADRLGILVIDEAPHVGTTRHYN-----PEFOKIA-----EDNIRMIIDR 368
 DB 344 LADREGLVVIDETPAVGV--HLNFMATTTGEGSERVSTWEKJRTPEHHODVIRELSR 400
 QY 369 HKNHPSVIMSVANBPESNHPDAGFPKALYETANEND-RTRPVVMSMDAPDERTDV 427
 DB 401 DKHNPSSVMSIANEATEEGAYEFKPYVELKRELDPOGRPTIYLFVWATEB--TDK 458
 QY 428 AKTFPDIVCNRYGYWYIYGRIBEGLOALEKDIIELYARRH---KPIFTEFGADAIA 483
 DB 459 VAEILDIVALNRVNGWYFDGDLBAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTV 515
 QY 484 GIHYDPQMESESEYQAEIVETIRLLKKDYIIGTHVAFADPFTPOVVRPILNHKGVF 543
 DB 516 GFHDIDPVMTPEEYQVEYYQANHVAFDEFENFVGEQAWNFADFATISQGMVRYQNGKGVF 575
 QY 544 TRDROPKLVAHVRRLMSEV 563
 DB 576 TRDRKPLAAHVFRERWTNI 595
 RESULT 7
 AAW93821
 ID AAW93821 standard; protein, 618 AA.
 XX
 XX AAW93821;
 AC
 XX 25-JUN-1999 (first entry)
 DT
 XX
 DE Bacillus sp. GUS protein.
 XX
 XX GUS; beta-glucuronidase; secreted; reporter molecule; marker;
 KM receptor molecule; diagnostic tool; transgene construction; plant;
 KM insect; cleavage; detoxification; glucuronide.
 XX
 OS Bacillus sp.
 XX
 XX WO9913085-A2.
 EN
 XX 18-MAR-1999.
 PD
 XX
 PE 09-SEP-1998; 98WO-US019217.
 PR
 XX 09-SEP-1997; 97US-0058263P.
 PA
 XX (CAMB-) CAMBIA BIOSYSTEMS LLC.
 PI Jefferson RA, Killian A, Keese PK;
 DR WPI; 1999-229241/19.
 DR N-PSDB; AAX23825.
 XX
 PT New isolated microbial beta-glucuronidase.
 XX
 PS Claim 5; Fig 4A-C; 76pp; English.
 XX
 CC This invention describes a novel secreted form of Bacillus sp. beta-
 CC glucuronidase (BogUS). The microbial BogUS polypeptide can be used as a
 CC reporter/effector molecule and as a diagnostic tool. The products of the
 CC invention can be used as markers for transgene constructions, e.g. in
 CC plants or insects. They can also be used for the cleavage and
 CC detoxification of glucuronides and to examine conjugation patterns of
 CC glucuronides
 CC
 SQ Sequence 618 AA;

Query Match 31.0%; Score 929; DB 2; Length 618;

Best Local Similarity 35.6%; Pred. No. 5e-67;

Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```

QY 1 MVRPQRKKRFFILLNGVNNLEV-----TSKDRPIAVGSMNE--QYODLC 44
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17 MLVPINTETRGVPLNGVNNFKLDYKGLEBKYESKLTDTISMAVSSYNDIGVTEIR 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 45 YEGSPFYKTTFFVPKLSQKHRLRYFAAVNTDCEVPLNEKNGENHIELPREVNTGK 104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 77 NHIGVYEREFVPAVKQRIYLRFGSATPKAIVVNGELVHEHKGFLPEAEIINNS 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 105 VKSGENLRVVENRKLKVGSPSKVPDSCGTHVGFSGS-----PPANDFEPY 154
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 137 LRQGNARVYAVNDNL-----DSTLPVGLYSEHNBEGKVIKRPDFNVA 186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 GIIRPIEFTDARILIDWDTSESEPEKKLGKVKVIESEBAVQENTIKLG--EEE 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 187 GLHRPVKITTPFYVEDISVTDENGR---TGTVYTVDFQ---GKAETVKSVDDE 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 213 KKIIRTSRFEGERFLENARFMSLEDPIYPLKVELEKDEYTLDI-----GRTISMD 267
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 240 GKVVASTEGSGNVEIRPNVILMBPLNTLYQIKVELVNDGLITDVSEPPGVRIVEVNDG 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 268 RLTVNGKPVFLKFGKHEEPVVLGQGFYPLMIKDFNLKMINANSFRTSHYPYSEWLD 327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 300 KFLINKPFYFKGFGKHEEDTPINGRGFNASNVMDFNILKIGANSFRTAHYPSSELMR 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 328 LADRLGLIVIDEAPHVIGITRYHN-----PEKOKIA-----EDNIRMIDR 368
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 360 LADREGLVIVIDEPAVG---HLNFMAVTGLGEGSERVSTWEKIRTEHHQDVLRELVS 416
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 369 HKHNPSTVMSVANPEPSNHPDAGFPKALYETANMD--RTPRVVMSMDADPERRDV 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 417 DKHNPSTVMSVANEATEEGAYEFKPLVELTKELDPQRPVTIYLVFMATPE--TDK 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 428 ALKXFDVVCNRYGYWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEFGADA 483
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 475 VAEIIVIAINRYGMVFDGDDLEAAKVHLR---QEFHANMKCPGKPIVITEGADTVA 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 484 GIHYDPQMFSESEYQAEIVKTRILLKKDYIIGTHVWAPADFKTPONVRRPILNHKGV 543
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 532 GFHDIDVMTFEEYQVEYQANHVFPDEFENFVEGQAMNFADPATSQVNRVQGNKKG 591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 544 TRDRQPLVAHVLRRLMSEV 563
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 592 TRDRKPLAAHVFRERWTNI 611
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8

ID AAB28408 standard; protein; 615 AA.

AC AAB28408;

DT 26-JAN-2001 (first entry)

DE Codon-optimised Staphylococcus GUS.

XX Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;

KM Pseudomonas; Staphylococcus; Thermococcus; transgenic plant; bioindicator;

KW transgenic insect; marker; glucuronide detoxification.

OS Staphylococcus sp.

PN MO200055333-A1.

XX 21-SEP-2000.

PF 16-MAR-2000; 2000MO-US007107.

PR 17-MAR-1999; 99US-00270957.

PA (CAMP-) CAMBIA BIOSYSTEMS LLC.

XX Jefferson RA, Mayer JB;

XX WPI; 2000-647075/62.

DR N-PSDB; AAA07938.

XX Novel microbial beta-glucuronidase genes and gene products used as

PT reporter/effector molecule, as diagnostic tool, in positive selection, to

PT target molecules to specific cells and to detect and track linked genes.

XX Claim 3; Fig 5B; 116pp; English.

The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS genes were obtained from six different genera: Enterobacter/Salmonella, Pseudomonas, Salmonella, Staphylococcus and Thermococcus. Microbial GUS can be used as a reporter/effector molecule for transgenic constructions and in vitro diagnostic applications. It may also be used to generate sentinel plants that serve as bioindicators of environmental status. It may be used to generate transgenic insects for tracking insect populations or to facilitate the development of a bioassay for compounds that affect molecules critical for insect development (e.g. juvenile hormone). Secreted GUS may also serve as a marker for beneficial fungi destined for release into the environment. In animal systems, secreted GUS may be used to achieve extracellular detoxification of glucuronides (e.g. toxin glucuronide) and to examine conjugation patterns of glucuronides. Microbial GUS may also be used in traditional medical diagnostic assays, for drug testing, pharmacokinetic studies, following bioavailability studies, diagnosis of diseases and syndromes, following progression of disease or its response to therapy. Microbial GUS has increased thermal stability, high turnover number and enzymatic activity. It is highly specific for the substrate and water soluble, and the substrates are stable

Sequence 615 AA;

Query Match 30.8%; Score 925; DB 3; Length 615;

Best Local Similarity 35.7%; Pred. No. 1.1e-66;

Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

```

QY 4 PQRKKRFFILLNGVNNLEV-----TSKDRPIAVGSMNE--QYODLCYEE 47
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 9 PINTERRGVADLNGVNNFKLDYKGLEBKYESKLTDTISMAVSSYNDIGVTEIRNHI 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 48 GFTYKTTFFVPKLSQKHRLRYFAAVNTDCEVPLNEKNGENHIELPREVNTGKVS 107
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 69 GYVYEREFVPAVKQRIYLRFGSATPKAIVVNGELVHEHKGFLPEAEIINSLRD 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 108 GENELRVVENRKLKVGSPSKVPDSCGTHVGFSGS-----PPANDFEPYGGII 157
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 129 GMRVYVAVNDNL-----DSTLPVGLYSEHNBEGKVIKRPDFNVAIGH 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 158 RPYLIEFTDARILIDWDTSESEPEKKLGKVKVIESEBAVQENTIKLG--EEBK 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 RPKVITTPFTYVEDISVTDENGR---TGTVYTVDFQ---GKAETVKSVDDEGKV 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 KTSNRVVEGEFLENARFMSLEDPIYPLKVELEKDEYTLDI-----GRTISMDER 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 232 VASTEGISGNVEIRPNVILMBPLNTLYQIKVELVNDGLITDVSEPPGVRIVEVND 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 271 LNKRPVFLKFGKHEEPVVLGQGFYPLMIKDFNLKMINANSFRTSHYPYSEWLD 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 292 INNKPFYFKGFGKHEEDTPINGRGFNASNVMDFNILKIGANSFRTAHYPSSELM 351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 331 RIGILVIDEAPHVIGITRYHN-----PEKOKIA-----EDNIRMIDRHN 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 352 REGLVIVIDEPAVG---HLNFMAVTGLGEGSERVSTWEKIRTEHHQDVLRELVS 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 HPSVIMSVANPEPSNHPDAGFPKALYETANMD--RTPRVVMSMDADPERDVALK 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 409 HPSVIMSVANPEATEEGAYEFKPLVELTKELDPQRPVTIYLVFMATPE--TDV 466
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 YFDIVVCNRYGYWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEFGAD 486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

D6      467  LIDVIALNRNGVYFPGGDI EAAKYHLR --OEFAMNKRCKCPKFMITEYGAULTVAGFH 522
QY      487  YDPQPMFSESYQALVELKTI RLKKDYI IGHVAFADFKTPQNVRRPILNHKGVTRD 546
D6      524  DIDPMVTEESYQVEYYQANHVPDEFENFVGQANMFADFATSGVMVGQNKGGVTRD 583
QY      547  RQPKLVANVLRRLMSEV 563
D6      584  RKPKLAAHVFERERTINI 600

```

RESULT 9
 AAB28409
 ID AAB28409 standard; protein; 602 AA.
 AC AAB28409;
 DT 26-JAN-2001 (first entry)
 XX
 DE *Salmonella* beta-glucuronidase.
 XX
 KW Microbial; beta-glucuronidase; GUS; Enterobacter; *Salmonella*;
 XX Pseudomonas; *Staphylococcus*; Thermotoga; transgenic plant; bioindicator;
 XX transgenic insect; marker; glucuronide detoxification.
 XX
 OS *Salmonella* sp.
 XX
 PN WO20005333-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000MO-US007107.
 XX
 PR 17-MAR-1999; 99US-00270957.
 XX
 PA (CAMP-) CAMBIA BIOSYSTEMS LLC.
 XX
 XX Jefferson RA, Mayer JE;
 XX
 DR WPI: 2000-647075/62.
 XX
 DR N-PSDB; AAA079339.
 XX
 PT Novel microbial beta-glucuronidase genes and gene products used as
 PT reporter/effector molecule, as diagnostic tool, in positive selection, to
 PT target molecules to specific cells and to detect and track linked genes.
 XX
 PS Example 3; Fig 17; 116bp; English.
 XX
 XX The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
 CC genes were obtained from six different genera: Enterobacter/*Salmonella*,
 CC Pseudomonas, *Salmonella*, *Staphylococcus* and *Thermotoga*. Microbial GUS can
 CC be used as a reporter/effector molecule for transgenic constructions and
 CC in in vitro diagnostic applications. It may also be used to generate
 CC sentinel plants that serve as bioindicators of environmental status. It
 CC may be used to generate transgenic insects for tracking insect
 CC populations or to facilitate the development of a bioassay for compounds
 CC that affect molecules critical for insect development (e.g. juvenile
 CC hormone). Secreted GUS may also serve as a marker for beneficial fungi
 CC destined for release into the environment. In animal systems, secreted
 CC GUS may be used to achieve extracellular detoxification of glucuronides
 CC (e.g. toxin glucuronide) and to examine conjugation patterns of
 CC glucuronides. Microbial GUS may also be used in traditional medical
 CC diagnostic assays, for drug testing, pharmacokinetic studies,
 CC bioavailability studies, diagnosis of diseases and syndromes, following
 CC progression of disease or its response to therapy. Microbial GUS has
 CC increased thermal stability, high turnover number and enzymatic activity.
 CC It is highly specific for the substrate and water soluble, and the
 CC substrates are stable
 XX
 XX Sequence 602 AA;

Best Local Similarity 35.3%; Pred. No. 2.6e-66;
Matches 213; Conservative 110; Mismatches 223;
Indels 58; Gaps 11;

```

0Y      1  MVRQDRNKRRIILLINGWMLNLEVYSKO-----RRIAYGSMNEY--ODL 43
0Y      1  MVRQDRNKRRIILLINGWMLNLEVYSKO-----RRIAYGSMNEY--ODL 43
Db      1  MLRSVETATREIKKLDGLMSFCMSDSEECGNAQQMWROBPLPQSRALIAVGSYNDQFAAAEI 60
0Y      44  CYEEGPFYKTTFFVFPVKCSOKHRLYPAAVNTDCEVFLNKGXGEMHIELPREVDYTG 103
0Y      61  RNYYGNWYQOEIRI PKGMWRQRVLPRFDAATHYXKVAWVDQFMHEHOGGTTPEADISH 120
0Y      104  KVXGGEN-ELRVVENRLKVGFSPSKVBDSGTHTVGFEFGSPPPANF--DFEYGGIIRPVL 161
Db      121  LISGESRITVVCNNELNMOTIP-----PQVTOGNGKKQQAIFYHDFEYAGIHSVM 175
0Y      162  IETDHAARIIDWDTSESEBEKKLKVXKXIEVSEAVGQEMTIKLGEEBKIRTSNRF 221
Db      176  LYTTPKTFVEDITVAVTOVAD---DLAQTVAMVRANG--EVRVELRDAEQQLVASQOG 229
0Y      222  VEGEFILENAPFSLDEPILYPLKY----ELEKREYTLIDIGIRITSMDEKLYLNGKRVF 277
Db      230  EKGELLLEGPRLMOPGEGYLTELRLVIAOHODEQDEPLRGVIRSEVXKGEQFLNNHPEY 289
0Y      278  LKGFGEKEEFPVLQCGTFYPLMIDQNLKIKANNSEFRTSHYPXYSEEWLADLADLGLIVI 337
Db      290  FTGGRHEDDALRKGRPDNYLMMYDHALMBDITGANSRTSHYPPAAEBLDMADDEHGIVII 349
0Y      338  DEADHVCI-----TRYHNPETOIAEDNIRRMIDRKHNPDSYIWM 379
Db      350  DETAAVGFNLISLIGSFVGEKPKELYDEAVNDETQRAHQAIELLARDGNHPSVWMS 409
0Y      380  VANPESNHPRAEAEFKALYETAEMEMRTRPVVWNSM--DAPERRDVALKXFDYVCV 437
Db      410  IANBPDTRENGAREYFAPRLAOTRELDPTREITVCVNWVFCABSDTTIDL---FDVVC 465
0Y      438  NRYVGMWYIQGRIEEGQALEKEKIDIEELYARHKEIFYTEFGADAIAGIHYDPOMFSEBX 497
Db      466  NRYVGMWYQSDLDKAEKVELEKELLANQEKLHRIIITTEYVDTLAGIHSYNNMMSSEY 525
0Y      498  QAEIVERTIRLLKKDYIIGTHWAFADFKTPQVNRRLPINHKGVFTFRDROPKLVANLVR 557
Db      526  QCAMLDIMYHFRDVSVAVVGQVWVNFADFAFQSGIMEVGNKKGIPTFRDRKPSAAFLQ 585
0Y      558  RLMS 561
0Y      586  KRW 589

```

RESULT 10	
AA048998	
ID	AA048998 standard; protein; 598 AA
XX	
AC	
XX	AA048998;
XX	
DT	10-MAY-2002 (first entry)
XX	
DE	L gasseri beta-glucuronidase.
XX	
KW	Beta-glucuronidase; GUS; low pH; L
XX	
OS	Lactobacillus gasseri.
XX	
PN	W0200190305-A2.
PD	
XX	
PD	29-NOV-2001.
XX	
PF	22-MAY-2001; 2001WO-US016667.
XX	
PR	23-MAY-2000; 2000US-0206372P.
XX	
PA	(UNNC-) UNIV NORTH CAROLINA STATE.
XX	
PI	Russell WM, Klaenhammer TR;

XX WPI, 2002-062529/08.
 DR N-PSDB; AAL44836.
 XX Novel isolated nucleic acid of Lactobacillus gasseri encoding beta
 PT glucuronidase having activity at acidic pH which is useful as reporter
 PT protein for highly aciduric organisms, or as marker of transformed cells.
 XX
 PS Claim 19, Page 49-51; 53pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC Lactobacillus gasseri beta-glucuronidase (GUS). The protein has maximum
 CC activity at acidic pHs, and thus can be used to detect low pH
 CC environments. The present sequence is the protein of the invention
 XX
 SQ Sequence 598 AA;
 Query Match 30.6%; Score 919.5; DB 5; Length 598;
 Best Local Similarity 34.1%; Pred. No. 2.9e-66;
 Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;
 QY 4 PQRKKRPFILINGVWNLVLT-----SKDRP-----IAGGSWNE--QYDLCYEE 47
 DB 7 PIQKRYRFTNLNMGTFQETDPNSVGLDEGWNKELPDEEMPVGTFAELTTKKRKYTT 66
 QY 48 GPFYKTFYVVKLSQKHRTLYFAAVNTDCEVFLNGEKYGEHIEYLPEVDTGKYKS 107
 DB 67 GDFWYQDFPFLPSFLKKELYIRGSVTHRAKVPINHEVGHGGLPQVKISNINY 126
 QY 108 GE-NELRVVENRLKVGFP---SKVDSGTHVGFSGFPFANFDFPFGIIRPLYIE 163
 DB 127 DQTRVTVVUNNELSEKAIKPGCTEILDNG-----KLAQPFDFPNFNGIRANWML 179
 QY 164 FTDHARLIDVWDTSESEPEKKLGKVKIIVSEAVGQ-EMTIKIGEEKIIRTSRFV 222
 DB 180 ALPQSQTNFELN-----YQLANNKATITVYNIENANNAEFKVTLPDNGQEVACATSKN 232
 QY 223 EGEFTLNARFMSLEDYVLYLKYELK-----DEYTLDGIRITSDERKLYNGKPVF 277
 DB 233 TSSLTITNPHLMSNDPYSYKIKIEMLEDGKTVEYDKIGIRVKLVNDKILNNHPIT 292
 QY 278 LKGGKHEEFPVLCQGTFFPLIMIDPFLMKWINANSFRTSHYPSEEMLDLADRLGILVI 337
 DB 293 LKGGKHEEFPVLCQGTFFPLIMIDPFLMKWINANSFRTSHYPSEEMLDLADRLGILVI 352
 QY 338 DEAPHVIGITRYHN-----PETOXIABDNIRMTDRHKNHPSVIMWS 379
 DB 353 DEVPAVGLNRSITNPLVNTSNQSHFPASKTVPFLKIKVHEQIEIKEMIDRDQRHPSVIAMS 412
 QY 380 VANPESNHPDABEGFALVETANEMD-RTRPVVWVGMMDAPDERTDVALKTYFDIVCN 438
 DB 413 LFNPESTTQESYDFDIFAFARKLPDQNPRTYGTLMWSGPK--VDKHLPLCDFYCLN 470
 QY 439 RYVGWVYVYG-RLEGQALEKEDIEELY-ARHRKPIFVTEFGADAIAGIHYPQWSE 496
 DB 471 RYVGWVYVYG-RLEGQALEKEDIEELY-ARHRKPIFVTEFGADAIAGIHYPQWSE 530
 QY 497 YQAEVLEKTRILLLKKYIITGTHWAPADFTPONVRPILNKGVTTRDQPKLVAVHL 556
 DB 531 YQAEVLEKTRILLLKKYIITGTHWAPADFTPONVRPILNKGVTTRDQPKLVAVHL 590
 QY 557 RRLWSEV 563
 DB 591 KKRWQQL 597
 RESULT 11
 ID ADF28924 standard; protein; 648 AA.
 XX ADF28924;
 XX 12-FEB-2004 (first entry)

XX DE Murine beta-glucuronidase precursor.
 XX KM Mouse; beta-glucuronidase; gene therapy; cytostatic; antidiabetic;
 XX immunosuppressive; hepatotropic; adeno-associated virus.
 XX OS Mus sp.
 XX FH Key Location/Qualifiers
 FT Peptide 1..22
 FT Protein /label= Signal_peptide
 FT /label= 23..648
 FT /label= Beta-glucuronidase
 XX W02003089011-A1.
 XX PD 30-OCT-2003.
 XX PF 21-APR-2003; 2003MO-US012324.
 XX PR 19-APR-2002; 2002US-0374083P.
 XX PA (UYFL) UNIV FLORIDA.
 XX PI Atkinson MA, Floete TR, Song S, Toiler SA;
 DR WPI, 2003-845502/78.
 PT New adeno-associated viral vector, useful in preparing a composition for
 PT treating or preventing e.g., cancer, diabetes, or autoimmune, pancreatic
 PT or liver disease.
 XX
 PS Example 4; SEQ ID NO 23; 183pp; English.
 XX
 CC The present sequence is that of murine beta-glucuronidase precursor (beta
 CC -D-glucuronoside glucuronosylhydrolase). This is an example of a
 CC therapeutic protein that can be encoded by an adeno-associated virus
 CC (AAV) vector of the invention. Such vectors comprise a promoter operably
 CC positioned upstream of a nucleic acid encoding a biologically active
 CC therapeutic mammalian serpin or cytokine polypeptide, and optionally also
 CC include an enhancer sequence and a post-transcriptional regulatory
 CC sequence. A recombinant AAV virion comprising the vector, and a mammalian
 CC cell (preferably an endothelial, islet, hepatocyte, pancreas, kidney,
 CC muscle, spleen, liver, heart, lung, or brain cell) comprising the vector
 CC are claimed. A claimed composition comprises the vector, the recombinant
 CC AAV virion, AAV viral particles, or the mammalian cell, and is used in
 CC cancer, diabetes, autoimmune disease, pancreatic disease or liver disease
 CC therapy. The composition is also used in claimed methods for preventing
 CC type I diabetes, and for reducing the rate of disease progression of type
 CC I diabetes, in a human.
 XX
 SQ Sequence 648 AA;
 Query Match 30.1%; Score 902.5; DB 7; Length 648;
 Best Local Similarity 35.7%; Pred. No. 7.9e-65;
 Matches 223; Conservative 103; Mismatches 207; Indels 91; Gaps 21;
 QY 1 MVRPQRKKRPFILINGVWNLVLT-----SKDRP-----IAGGSWNEQ 39
 DB 27 MVRPQRKKRPFILINGVWNLVLT-----SKDRP-----IAGGSWNEQ 86
 QY 40 YODLCYEE--GPFYKTFYVVKLSQ---KHRLTFAAVNTDCEVFLNGEKYGEHIEY 94
 DB 87 YQAEVLEKTRILLLKKYIITGTHWAPADFTPONVRPILNKGVTTRDQPKLVAVHL 146
 QY 95 LPEFVDTGKVGSG--ENELRVVENRLKVGFPSPKVP-----DSGTHVGFSGP 143
 DB 147 LPEFADISKLKVGSGPLTTCRITAINNTLT-----PHTLPRTGVYKTTDMSYKGYF--V 200
 QY 144 PPRANDFPFGYGIIRPLYIEFTDHAIRLIDVWDTSESEPEKKLGKVK--VKIEVSEAVG 201
 DB 201 QDTSFDFPNVAGIHRGAVLTPPTTYIDITVITN---VEQDIGLVYMWISVQSGSEHF-- 255

```
OY 202 QEMTIKLGEEBKIRTSNRFEVEGEFILENARFW----SLEDP-YLYPLKV-----ELEKD 251
DB 256 -QLEVOQLDEGCKVVAHGNGOQLQVPSANIMWYLMHEHPAYMSLEVKVTTTESYTD 314
OY 252 EYTLDIGIRITISWDEKRLYLNGKRPVFLKGFGESEFPVLGGTFYPLMIKDPNLKWTNA 311
DB 315 YTTLPVIGIRTVAVTKSKFLINGKPPFQGVNKHGSDIRGKFDWPLVKDFNLRLWIGA 374
OY 312 NSFRTSHYPSYSEEMDLADRGLIYLIDEAPHVIGIT-----RYHNPETOKIADN 361
DB 375 NSFRTSHYPSYSEEVQLCDRYGIIVIDECPGVGIVLPQSPGNSLRHNL-----EVMBEL 429
OY 362 IRRMIDRKHNPSPVIMSVANEPESNHPDAGEFFKALYETANEMDRTPRVVWS--MMDA 419
DB 430 VR----RDKNHPAVVMSVANEPSSALKRAAYFFKTLITHKALDLTRPVTVSNAKTDA 485
OY 420 PDETRDVALKYPDIVCNRYGWTIYQGRIBEGQLAEKDI EELIYAHRKPIFYTERGA 479
DB 486 -----DIGAPYVDVICNVSYSFWDYGHLEVIQPOLNSQFENWYKTHQKIIQSEYGA 539
OY 480 DALAGIHDPPEMSESEYQAELEVETIRLL--LKQDYIIGTHVMAFADPKTPQVNRRL 537
DB 540 DALPEIHEDPPRMSESEYQAKAVLENYHSVLQDKRKEXYVGGELIMNFAFMTOQSLRVI 599
OY 538 NHKGVETRDROPKLVAVHVLRLW 560
DB 600 NKGIFTRQRPKTSAFILRERYW 623
```

```
RESULT 12
AAE02444
ID AAE02444 standard, protein, 648 AA.
```

```
AC AAE02444;
DT 10-AUG-2001 (first entry)
XX
DE Murine beta-glucuronidase (GUS).
XX
KM Murine; adeno-associated viral expression vector; AAV; gene therapy;
KW lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
KW Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Peptide 1..26
FT Protein /label=Signal_peptide
FT 27..648
FT /note="Murine mature beta-glucuronidase (GUS)"
XX
PN MO200136603-A2.
PD 25-MAY-2001.
XX
PE 17-NOV-2000; 2000MO-US031688.
XX
PR 17-NOV-1999; 99US-0166097P.
PR 30-JUN-2000; 2000US-0215430P.
XX
PA (AVIG-) AVIGEN INC.
PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
P1 Podsakoff G, Watson G, Couto LB, Yang B;
XX
XX WPI; 2001-343814/36.
DR N-PSDB; AAD06387.
XX
XX Use of recombinant adeno-associated virus, comprising gene encoding a
FT protein defective or missing in lysosomal storage disease, in the
FT manufacture of a medicament for treating the lysosomal storage disease.
XX
PS Disclosure; Page 85-87; 97pp; English.
```

```
XX
CC The present invention relates to recombinant adeno-associated virus (AAV)
CC expression vectors and virions, which include genes coding for enzymes
CC defective or missing in lysosomal storage disease (LSD). AAV is useful in
CC the manufacture of a medicament for treating lysosomal storage disease
CC e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome) is due
CC to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS) which
CC aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in gene
CC therapy. The present sequence is murine GUS protein. This sequence is
CC used in AAV constructs
XX
SQ Sequence 648 AA;
```

```
Query Match 30.0%; Score 899.5; DB 4; Length 648;
Best Local Similarity 35.9%; Pred. No. 1,46-64;
Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;
```

```
OY 1 MVRPQNRKKRFLILNGVNLLEV--TSKDR-----PIAVPSMNEQ 39
DB 27 MLFPKESREIKALDGLMHRADLSNNRLOGEQWTRQPLRSGPVLDMVPSFNDI 86
OY 40 YQDLCEE--GPFTYKTFYVPKLSQ--KHIRLYFAAVNTDCEVFLNGEKVGENHLEY 94
DB 87 TOEALRDFIGWVYERBALPRRWTDQDMKRVLRINSAHYVAVWVNGIHVEHGGH 146
OY 95 LPFVDVTKYKSG--ENELRVVENRLKYGGFPSKRP-----DSGHTTVGFFGSF 143
DB 147 LPFVADIKLVQSGELTTCRTITAIINTLT-----PHTLPQGITVYKKTJSMVPKGYF--V 200
OY 144 PRANDFPPYGIIRPVLIIEFTDRIIDWDTSESEPEKLGKVK--VKRIESEAAG 201
DB 201 QDTSDFDNVAGLHRSVLYTPTTYIDIDIVITN--VEQDIGLYTWISVQSEHF-- 255
OY 202 QEMTIKLGEEBKIRTSNRFEVEGEFILENARFW----SLEDP-YLYPLKV-----ELEKD 251
DB 256 -QLEVOQLDEGCKVVAHGNGOQLQVPSANIMWYLMHEHPAYMSLEVKVTTTESYTD 314
OY 252 EYTLDIGIRITISWDEKRLYLNGKRPVFLKGFGESEFPVLGGTFYPLMIKDPNLKWTNA 311
DB 315 YTTLPVIGIRTVAVTKSKFLINGKPPFQGVNKHGSDIRGKFDWPLVKDFNLRLWIGA 374
OY 312 NSFRTSHYPSYSEEMDLADRGLIYLIDEAPHVIGIT-----RYHNPETOKIADN 361
DB 375 NSFRTSHYPSYSEEVQLCDRYGIIVIDECPGVGIVLPQSPGNSLRHNL-----EVMBEL 429
OY 362 IRRMIDRKHNPSPVIMSVANEPESNHPDAGEFFKALYETANEMDRTPRVVWS--MMDA 419
DB 430 VR----RDKNHPAVVMSVANEPSSALKRAAYFFKTLITHKALDLTRPVTVSNAKTDA 485
OY 420 PDETRDVALKYPDIVCNRYGWTIYQGRIBEGQLAEKDI EELIYAHRKPIFYTERGA 479
DB 486 -----DIGAPYVDVICNVSYSFWDYGHLEVIQPOLNSQFENWYKTHQKIIQSEYGA 539
OY 480 DALAGIHDPPEMSESEYQAELEVETIRLL--LKQDYIIGTHVMAFADPKTPQVNRRL 537
DB 540 DALPEIHEDPPRMSESEYQAKAVLENYHSVLQDKRKEXYVGGELIMNFAFMTOQSLRVI 599
OY 538 NHKGVETRDROPKLVAVHVLRLW 560
DB 600 NKGIFTRQRPKTSAFILRERYW 623

RESULT 13
ABB70164
ID ABB70164 standard, protein, 643 AA.
XX
XX ABB70164;
XX
XX 26-MAR-2002 (first entry)
DT Drosophila melanogaster polypeptide SEQ ID NO 37284.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
```

Query Match	29.9%	Score 897	DB 4	Length 643
Best Local Similarity	34.2%	Pred. No. 2.2e-64		
Matches 216	Conservative 122	Mismatches 205	Indels 88	Gaps 20
QY	1	MVRPNRKKRFFLLINGVWNL-----EYTSKDR--PIAVPGSNMEO	39	
Db	1	MLYPRSESTRREVRSJDGIMNFVRSQANPTQGVRDWEYAKELSKSRPTTIPMPVPASVNDI	60	
QY	40	YQD-LCYEEGPTTYKTFYVPKLSQ-KHIRLYFAVWTDCEVFNNGEKVGENHLEYLPF	97	
Db	61	TTDNLRDHVGTVWYDRKFFVPRSMKQRIWLPFSVHYEAVWVINGKVVVHGHGHPF	120	
QY	98	EVDVYTKAKSG-ENELKRVVENRLKVGSP---SKVDSGCHTYVGFSGSPFPPANFDFFP	152	
Db	121	EAETVDTLSYAEENKRTIWCNALIQTTVPQGRITVEVPDGGMTI-----VQSYTFDEFN	175	
QY	153	YGGIIRPLIEFTTHARILIDIVDSESEPEKLGKVKIKLIESEAVGQ-----EMTK	207	
Db	176	YAGIHRSVHLTTPPTFIEEVEVTTNLSK-DATVEEIVYSVNSAANEADNVLTQIQAN	234	
QY	208	LGESEKTI--RTSNRFVEGEFFLENAFV-----SLEDPIYLYPKVELEK-----DEY	253	
Db	235	LYDDGILVIANATSDQKLGKLOVPVFWPMPYLMHSEPGYLYQLEIKILTNTDELIDVY	294	
QY	254	TLIDIGIRIISDEKRLYNGKRVFLKRGKKEEFPVLQGGFFYPLMIDPFLNLIKINNS	313	
Db	295	RLKKGIRTLSSNSQOFLINGKPYVFRGGRHEDSDIRGKGLDNALMVDPFNLKMTGANA	354	
QY	314	FRISHYPSSEWMLDLADRLGILVIDEADHVGITRYHNPETOK-----IA-----	358	
Db	355	YRTSHYPSSESMQFADHGIMIIDCEPSV-TELSTYS-DTSKGFLLMLASIFRNFQEO	412	
QY	359	-----EONIRIMIDIRKHQHPVIMVSVANEDESNHPRDAGEGFKALYETANEMDKTRPPVM	413	

Db	413	LLGKKSSLEQIHRDRNPSVVMMSIANEPRTGSVADSYPELVANFRSLDKTPI	472
Qy	414	USMNDAPBERTDVALKTYDVCARRYWYTYOGRIEGLQALEOIEBLA---	RHRK 470
Db	473	AIAY----SNTQDKGRSLDIISFRRYAMWYNAGILD--MITONVIDEALIMNRKYNK	525
Qy	471	PIFVEFGADAIAGIHYDPEPOMFSEEOALVEKTRIL---LLKKYIIGTHWAFADPK	527
Db	526	PIMSEYGDILEGILHMGPAYWSEEPQEVSRHFKADELRKKGMFIFGEFVWNPADPK	585
Qy	528	TPQNVRRPILNKKGVFTDRQPKLVAAHVLR	558
Db	586	TAQSYTRVGNNKGGVETRAPQPKAAHLRLK	616
RESULT 14			
ID	ADD45479	standard; protein; 648 AA.	
XX	ADD45479;		
AC	ADD45479;		
DT	29-JAN-2004	(first entry)	
DE	Rat Protein P06760, SEQ ID NO 10912.		
KW	Rac; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;		
KM	chronic constriction injury; CCI, spared nerve injury; SN1; Chung.		
OS	Rattus norvegicus.		
XX	WO2003016475-A2.		
PN	27-FEB-2003.		
PD	14-AUG-2002; 2002WO-US025765.		
XX	14-AUG-2001; 2001US-0312147P.		
PR	01-NOV-2001; 2001US-0346382P.		
FR	26-NOV-2001; 2001US-0333347P.		
XX	(GENO) GEN HOSPITAL CORP.		
PA	(PARB) BAYER AG.		
XX	Woolf C, D'urso D, Befort K, Costigan M;		
PL	WPI; 2003-266312/26.		
DR	GENBANK; P06760.		
PT	New composition comprising two or more isolated polypeptides, useful for		
XX	preparing a medicament for treating pain in an animal.		
PS	Claim 1; Page; 1017p; English.		
XX	The invention discloses a composition comprising two or more isolated rat		
CC	or human polynucleotides or a polynucleotide which represents a fragment,		
CC	derivative or allelic variation of the nucleic acid sequence. Also		
CC	claimed are a vector comprising the novel polynucleotide, a host cell		
CC	comprising the vector, a method for identifying a nucleotide sequence		
CC	which is differentially regulated in an animal subjected to pain and a		
CC	kit to perform the method, an array, a method for identifying an agent		
CC	that increases or decreases the expression of the polynucleotide sequence		
CC	that is differentially expressed in neuronal tissue of a first animal		
CC	subjected to pain, a method for identifying a compound which regulates		
CC	the expression of a polynucleotide sequence which is differentially		
CC	expressed in an animal subjected to pain, a method for identifying a		
CC	compound that regulates the activity of one or more of the		
CC	polynucleotides, a method for producing a pharmaceutical composition, a		
CC	method for identifying a compound or small molecule that regulates the		
CC	activity in an animal of one or more of the polypeptides given in the		
CC	specification, a method for identifying a compound useful in treating		
CC	pain and a pharmaceutical composition comprising the one or more		
CC	polypeptides or their antibodies. The polynucleotide or the compound that		
CC	modulates its activity is useful for preparing a medicament for treating		

path (e.g. spinal segmental nerve injury (Chung)) chronic constriction injury (CCI) and spared nerve injury (SNI) in an animal (e.g. gene therapy). The sequence presented is a rat protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at http://wipo.int/pub/publ/published_pcr_sequences.

SD Sequence 648 AA;

Query Match	29.8%	Score 894.5;	DB 7;	Length 648;
-------------	-------	--------------	-------	-------------

Best Local Similarity	36.1%;	Pred. No. 3,6e-64;	
Matches 225;	Conservative 98;	Mismatches 210;	Indels 91; Gaps 20

Qy	1	MVRPDRNKKRFLFLINGVWN-----LEVTSKORPIAVPSSNME	38
Db	27	MLPFRPTSRREKLVDGLMSFRADYSNNRLQSGFQKOWTRQPLREBGPFLIDMVPSSPNDI	86
Qy	39	-QYODLCSEBGRPTKTTFFVYVKLSOKHRIYLFAVAVTD---CEVFLNGEKVGNHLEY	94
Db	87	TOEALERNFICGWVYEREAVALPQKWTQDTDRVLRVLRINSAHYAAVVMVNGIHVHEGGH	146
Qy	95	LPFEVDYTGKTKSG--ENELRVVVENRLKVGFFSKPV-----DSGHTTVGFFGSF	143
Db	147	LPFEADITKLVSQSGELTTFRRVLTIANNTLT---SYTLPRGTIVYKTPESMYPKYF--V	200
Qy	144	PRANFDFPYGGIIRPVLIEFTDHAIRLIDVIDWTSSESEPEKKLGKVKYKIEVSEBAVQE	203
Db	201	QDISDFFNAYGNIHNSVLYTPTTYIDITYTT---DVDKRVGLVNWISV-QGSDHFQ	256
Qy	204	MTIKLGESEKKIKRTSNRFEVGEFFILENAPFW---SLEDP-YLYPLKY----ELEKDEY	253
Db	257	LEVRLLEDGKIVARGTNEGOLKVPRAHLMWRYLHNEHPALYLSLEVMTPTPESVSDPY	316
Qy	254	TLDGIRFISDMEKRLYNGKRVPLKFGKGEFFVLQSGTFEYPLMIDPNILKKTINNS	313
Db	317	TLPVGIRVVAATKSKFLINGKRFYFQGVNKGHSDSDIRGFGDWPILLIDPFNLRLVGNNS	376
Qy	314	FRTSHYPSSEBMLDLADRLGLVIDEARPHVGIT-----RYHNPETOKIAEDNIR	363
Db	377	FRTSHYPSSEBVLQICDRYGVIVIDECHQGVGLVLPQSGVNSLRNHL-----EVMDELVR	431
Qy	364	RMIDRHKNHPSVIMSVANBEPESNRPDAEGFPKLYETANEMDRFRPVLVMSAMDAPDER	423
Db	432	---RDKNHPVVMVMSVANBEPVSSILKRPAGYFFKTLIAHTKALDPPRPPTVPS-----N	480
Qy	424	TR---DVALKRFEDIQVNRYYGMYVYQGRIEBGLQALEKDIIEELYARRKRPFTVEFGAD	480
Db	481	TRYADMGARVVDVICVNSYLSMWDYGHLEIVIQGLQLSQPEWYKMYOKPIIGSEYCAD	540
Qy	481	AIAGIHYPDPOMFSEEQALVELKTIKILL--KKDYIIGTHVMAFADFKTPQNVKRPIL	537
Db	541	AVSGIHEBPPMPFSEEQYOTALLE--NYHLILDEKREKEYVIGELIWNFADFMNTQSLRVTG	599
Qy	538	NHKGVFTDRQPKLVAVHLR-RLW 560	
Db	600	NKKGIFTKRNPKMAFAILREKRYW 623	
RESULT 15			
ADES7446			
ID	ADE57446	standard; protein, 648 AA.	
XX	AC	ADE57446;	
XX	DT	29-JAN-2004 (first entry)	
XX	DE	Rat Protein P06760, SEQ ID NO 3307.	
XX	KN	Rat; pain, neuronal tissue; gene therapy; spinal segmental nerve injury;	
XX	KS	chronic constriction injury; CCI; spared nerve injury; SNI; Chung.	
XX		Rattus norvegicus.	

XX WO2003016475-A2.
PN
XX
XX 27-FEB-2003.
PD
XX
XX 14-AUG-2002; 2002WO-US025765.
PF
XX
XX 14-AUG-2001; 2001US-0312147P.
PR
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
PI
XX
XX WPI; 2003-268312/26.
DR
DR GENBANK; P06760.
XX
XX
PT New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
PS

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polynucleotides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antihoddes. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constiction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a rat protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at http://wipo.int/pub/published_pct_sequences.

SQ Sequence 648 AA;

Query Match	29.8%	Score 894.5	DB 7	Length 648
-------------	-------	-------------	------	------------

Best Local Similarity 36.1%; Pred. No. 3.6e-64;
Matches 225; Conservative 98; Mismatches 210; Indels 91; Gaps 20

Qy	1	MVRPQNRKKRFLLLINGVNN-----	-LEVTSKDRPIAVPGSMNE-	38
		: : : : : : : : : : : :	: : : : : : : : :	
Db	27	MLPFPKETSREBLKVLVDGLMSFPADYSNNRLQGFQEKOWYRQPLRESGPPLTMDMPVPSFNDI		86
Qy	39	-QYODLCEBEGFPYTKFTTFYVFPKLSOKHRIYPAAVNPD-----	CEVFLNGEKKGNHIEY	94
		: : : : : : : : : : : :	: : : : : : : : :	
Db	87	TOEAEIRNFICGWWYTEREAVLDPQRWTDOTDRVVLRLNSAHYAAVAVWVGIIHVEEGGH		146
Qy	95	LPFEDVLTGKYSG--ENELRVVENLRKLVQGFSPKVP-----	DSGTHIVGFGSH	143
		: : : : : : : : : : : :	: : : : : : : : :	
Db	147	LPFEADIRKLVQSGPLTFPRVLTAINNTLL-----	PYTLPRGIIYVTKTDSMPKGF--	V 200
Qy	144	PPANDFPFGGIIIRPVLIEFTDARIIDVWDTSESEPEKKLGKVKKIVIEVSEAVQGE		203
		: : : : : : : : : : : :	: : : : : : : : :	
Db	201	QDISDFENYAGLHNSVVLVYTPPTIYIDITVTT-----	DVDRVGLAVNMISV-QGSDHFG	256

Qy 204 MTIKIGEEBKIRTSNRPVEGSEFIENARFW----SLEDP-YLYPLKV-----ELEKDEY 253
Db 257 LEVRLLEDGKIVARGTNEGQLKVPRAHLMWPYLMHSPALYSLVYMTTPEBSVDFY 316
Qy 254 TLIDIGRTISMDKRLYLNGKPFVFLGKHEEPVLGQTFYPLMIKOPNLKWINANS 313
Db 317 TLPVGIRTVAVTKSKFLNGKPFYFQGVNKGHEDSDIRGRGPDWPLIKDFNLRLGANS 376
Qy 314 FRTSHYPYSEEWLADRLGILVIDEAPVVGIT-----RYHNPETQKIAEDNIR 363
Db 377 FRTSHYPYSEEVLOCDRYGIIVIDECPGVGIVLPQSPGNVSLRHHL-----EYWDDELVR 431
Qy 364 RMIDHKHPSVYIMSVANEPESNHPDAEGPEKALYETANEMDRTRPVVMTSMMDAPDER 423
Db 432 ----RDKNHPAVVMSVANEPVSSLKPPAGYIFKTLIAHTKALDPTRPVTFVS-----N 480
Qy 424 TR--DVALKYFDIVCVNRYGMYIYQGRIEGLQALEKDIIELYARHKPIFYTEFGAD 480
Db 481 TRYDADMGAPYVDVICVNSYLSWYHDYGHLEVIQQLTSQFENWYKMYQKPIIQSEYCAD 540
Qy 481 AIAGITHDPQMFSESEYQAEIYEKTIKILL--KKDYIIGTHWAFADFKTPQNVRRPIL 537
Db 541 AVSGLHEDPPRPFSESEYQJALLE-NYHLILDEKREKYEYIGELIWNFADFMTNQSPLRYTG 599
Qy 538 NHKGVFTRDROPKLVAHVLR-RLW 560
Db 600 NKGIFTROKNPKMAAFILRERYW 623

Search completed: January 24, 2005, 07:50:46
Job time : 92 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 07:33:21 ; Search time 26 Seconds
(without alignments)
1436.040 Million cell updates/sec

Title: US-09-936-759-6
Perfect score: 3001
Sequence: 1 MVRPQNNKKRFLILNGVWN.....TRDRPKLVAVHRLNSEV 563

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	4	US-09-270-957-6
2	2999	99.9	563	4	US-09-270-957-21
3	929	31.0	602	3	US-09-149-727-2
4	929	31.0	602	4	US-09-270-957-2
5	929	31.0	602	4	US-09-270-957-8
6	929	31.0	602	4	US-09-270-957-15
7	929	31.0	618	3	US-09-149-727-4
8	925	30.8	607	3	US-09-149-727-8
9	925	30.8	607	4	US-09-270-957-18
10	925	30.8	615	4	US-09-270-957-28
11	919.5	30.6	598	4	US-09-862-660-2
12	899.5	30.0	648	4	US-09-715-858-4
13	882	29.4	1010	4	US-09-118-276-12
14	879	29.3	603	3	US-09-149-727-6
15	879	29.3	603	4	US-09-270-957-17
16	877	29.3	603	4	US-09-270-957-23
17	877	29.2	832	3	US-08-630-820-7
18	877	29.2	832	4	US-09-273-453-7
19	875	29.2	604	4	US-09-893-525-37
20	875	29.2	659	4	US-09-893-525-40
21	875	29.2	850	4	US-09-893-525-42
22	872.5	29.1	602	4	US-08-882-704A-5
23	872.5	29.1	602	4	US-09-151-957-5
24	872.5	29.1	602	4	US-10-195-158-5
25	872.5	29.1	602	6	5432081-2
26	872.5	29.1	1242	4	US-09-488-270A-2
27	865.5	28.8	613	3	US-09-149-727-5

28	865.5	28.8	613	4	US-09-270-957-16	Sequence 16, Appl
29	865.5	28.8	651	4	US-09-715-858-2	Sequence 2, Appl
30	849.5	28.3	600	6	5268463-2	Patent No. 5268463
31	692.5	23.1	376	4	US-09-270-957-4	Sequence 4, Appl
32	692.5	23.1	376	4	US-09-270-957-19	Sequence 19, Appl
33	574	19.1	372	4	US-09-270-957-3	Sequence 3, Appl
34	574	19.1	372	4	US-09-270-957-22	Sequence 22, Appl
35	496	16.5	540	4	US-09-270-957-5	Sequence 5, Appl
36	489.5	16.3	535	4	US-09-270-957-20	Sequence 20, Appl
37	416	13.9	1334	6	5476657-1	Patent No. 5476657
38	414	13.8	1010	4	US-09-654-449-2	Sequence 2, Appl
39	414	13.8	1121	1	US-07-789-915A-2	Sequence 2, Appl
40	414	13.8	1121	1	US-08-005-002C-2	Sequence 2, Appl
41	414	13.8	1121	1	US-08-487-203A-2	Sequence 2, Appl
42	414	13.8	1403	1	US-07-908-253-3	Sequence 3, Appl
43	414	13.8	1403	2	US-08-694-865-17	Sequence 17, Appl
44	414	13.8	1403	2	US-08-535-837-3	Sequence 3, Appl
45	414	13.8	1403	3	US-09-124-491-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-270-957-6
Sequence 6, Application US/09270957
Patent No. 6641996
GENERAL INFORMATION:
APPLICANT: Richard A. Jefferson and Jorge E. Mayer
TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
FILE REFERENCE: 190106.405C1
CURRENT APPLICATION NUMBER: US/09/270,957
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 112
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 563
TYPE: PRT
ORGANISM: Thermotoga maritima
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(563)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-270-957-6

Query Match	99.9%; Score 2999; DB 4; Length 563;
Best Local Similarity	100.0%; Pred. No. 3e-256;
Matches	563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MVRPQNNKKRFLILNGVWNLEVTISKDRPIAVPGSMNQYODLCYEEGFTYKTFYVPK 60
DB	1 MVRPQNNKKRFLILNGVWNLEVTISKDRPIAVPGSMNQYODLCYEEGFTYKTFYVPK 60
QY	61 XLSQGHILYPAVAVTDCBFLNGSKVGENHLEYLPFVDVYGVKXSGENELRVAVENRL 120
DB	61 XLSQGHILYPAVAVTDCBFLNGSKVGENHLEYLPFVDVYGVKXSGENELRVAVENRL 120
QY	121 KYGSPSKVPDSCGTTGTFPGSFPFANPDPFYGGIIRPVLEFTDHRILDIWDTSES 180
DB	121 KYGSPSKVPDSCGTTGTFPGSFPFANPDPFYGGIIRPVLEFTDHRILDIWDTSES 180
QY	181 EPEKLVGKVKYKIEVSEBAVQCEMTIKLGESEKIRTSNRPVEGSEFIIENARFMSLEDPY 240
DB	181 EPEKLVGKVKYKIEVSEBAVQCEMTIKLGESEKIRTSNRPVEGSEFIIENARFMSLEDPY 240
QY	241 LYLPLVELKEQVTLIDIGIRITSMDEKRLVYNGKRVPLKGGKHEEPVLGGCTFYPLMI 300
DB	241 LYLPLVELKEQVTLIDIGIRITSMDEKRLVYNGKRVPLKGGKHEEPVLGGCTFYPLMI 300
QY	301 KDFNLKVINNSFRTSHYPSSEEWLADLADRLGILVIDEAPHVIGITRYHNPETOKIAD 360
DB	301 KDFNLKVINNSFRTSHYPSSEEWLADLADRLGILVIDEAPHVIGITRYHNPETOKIAD 360

```

QY 361 NRRMIDHKKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAP 420
DB 361 NRRMIDHKKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAP 420
QY 421 DERTDVALKYFDICVNRYYGWIYQGRIBEGLOALEKDIIEELYARHRKEIFYTEFGAD 480
DB 421 DERTDVALKYFDICVNRYYGWIYQGRIBEGLOALEKDIIEELYARHRKEIFYTEFGAD 480
QY 481 AAGIHDPQPMSESEYQAEIVKERTIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
DB 481 AAGIHDPQPMSESEYQAEIVKERTIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
QY 541 GVFTDRQPKLVAVHVRRLMSEV 563
DB 541 GVFTDRQPKLVAVHVRRLMSEV 563

```

RESULT 2
US-09-270-957-21

```

; Sequence 21, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270.957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermocoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-270-957-21

```

Query Match 99.9%; Score 2999; DB 4; Length 563;
Best Local Similarity 100.0%; Pred. No. 3e-256;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MVRPQRKKRFFILNGVWNLVTSKDRPIAVPSMNOYODLCYEEGPFYKTFYVVK 60
DB 1 MVRPQRKKRFFILNGVWNLVTSKDRPIAVPSMNOYODLCYEEGPFYKTFYVVK 60
QY 61 XLSQKHRLVPAVNTDCEVFLNGEKVGENHIEVLPFVDVTGKVGSGENELRVVENVRL 120
DB 61 XLSQKHRLVPAVNTDCEVFLNGEKVGENHIEVLPFVDVTGKVGSGENELRVVENVRL 120
QY 121 KVGFPSPKVPDSGHTVGFSGFPFPAVDFPPYGGIIRPVLIETDARLIDIVDTSES 180
DB 121 KVGFPSPKVPDSGHTVGFSGFPFPAVDFPPYGGIIRPVLIETDARLIDIVDTSES 180
QY 121 KVGFPSPKVPDSGHTVGFSGFPFPAVDFPPYGGIIRPVLIETDARLIDIVDTSES 180
DB 121 KVGFPSPKVPDSGHTVGFSGFPFPAVDFPPYGGIIRPVLIETDARLIDIVDTSES 180
QY 181 EPEKGLGVKVKIVSEEAQVQEMTIKLGESEKIRTSNRFVSEGFLENARFWSLEDPY 240
DB 181 EPEKGLGVKVKIVSEEAQVQEMTIKLGESEKIRTSNRFVSEGFLENARFWSLEDPY 240
QY 241 LYLPLKVELEKDEYTLIDIGIRTIISWDEKRLYLNGKPVFLKFGKIEEPVVGQGTFFYLM 300
DB 241 LYLPLKVELEKDEYTLIDIGIRTIISWDEKRLYLNGKPVFLKFGKIEEPVVGQGTFFYLM 300
QY 301 KDFPLKLVIANNSFRSHYPSSEEMDLADRLGLIVIDEAPHVIGITRYHNPETOKIAED 360
DB 301 KDFPLKLVIANNSFRSHYPSSEEMDLADRLGLIVIDEAPHVIGITRYHNPETOKIAED 360
QY 361 NRRMIDHKKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAP 420
DB 361 NRRMIDHKKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAP 420

```

```

QY 421 DERTDVALKYFDICVNRYYGWIYQGRIBEGLOALEKDIIEELYARHRKEIFYTEFGAD 480
DB 421 DERTDVALKYFDICVNRYYGWIYQGRIBEGLOALEKDIIEELYARHRKEIFYTEFGAD 480
QY 481 AAGIHDPQPMSESEYQAEIVKERTIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
DB 481 AAGIHDPQPMSESEYQAEIVKERTIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
QY 541 GVFTDRQPKLVAVHVRRLMSEV 563
DB 541 GVFTDRQPKLVAVHVRRLMSEV 563

```

RESULT 3
US-09-149-727-2

```

; Sequence 2, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Keesee, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA - GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149.727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-149-727-2

```

Query Match 31.0%; Score 929; DB 3; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.2e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```

QY 1 MVRPQRKKRFFILNGVWNLVTSKDRPIAVPSMNOYODLCYEEGPFYKTFYVVK 44
DB 1 MVRPQRKKRFFILNGVWNLVTSKDRPIAVPSMNOYODLCYEEGPFYKTFYVVK 44
QY 45 YEEGPFYKTFYVVKXLSQKHRLVPAVNTDCEVFLNGEKVGENHIEVLPFVDVTGK 104
DB 45 YEEGPFYKTFYVVKXLSQKHRLVPAVNTDCEVFLNGEKVGENHIEVLPFVDVTGK 104
QY 61 NHIGYVWEREFVPAVLDQRIVLFSGATHKALVYVNGELVHGKGFLEPEAEINNS 120
DB 61 NHIGYVWEREFVPAVLDQRIVLFSGATHKALVYVNGELVHGKGFLEPEAEINNS 120
QY 105 VKSGENELRVVENVRLKVGFPSPKVPDSGHTVGFSGS-----FPPANFDFPPYG 154
DB 105 VKSGENELRVVENVRLKVGFPSPKVPDSGHTVGFSGS-----FPPANFDFPPYG 154
QY 121 LRQGMKRVVAVNDNL-----DSTLTVGLYSEHSEGLKVTINRKRPDFFNVA 170
DB 121 LRQGMKRVVAVNDNL-----DSTLTVGLYSEHSEGLKVTINRKRPDFFNVA 170
QY 155 GIIIRPVLIETDARLIDIVDTSESSEPEKLGKVKKIVSEEAQVQEMTIKLG--EE 212
DB 155 GIIIRPVLIETDARLIDIVDTSESSEPEKLGKVKKIVSEEAQVQEMTIKLG--EE 212
QY 171 GIIIRPVLIETDARLIDIVDTSESSEPEKLGKVKKIVSEEAQVQEMTIKLG--EE 223
DB 171 GIIIRPVLIETDARLIDIVDTSESSEPEKLGKVKKIVSEEAQVQEMTIKLG--EE 223
QY 213 KIIIRTSNRFVSEGFLENARFWSLEDPYLPLKVELEKDEYTLIDIGIRTIISWDEK 267
DB 213 KIIIRTSNRFVSEGFLENARFWSLEDPYLPLKVELEKDEYTLIDIGIRTIISWDEK 267
QY 224 GKVAVSTEGSGVTELVNVLWBPVLTLYQIKVELVNGLTIDVYEEBPGVATVENVG 283
DB 224 GKVAVSTEGSGVTELVNVLWBPVLTLYQIKVELVNGLTIDVYEEBPGVATVENVG 283
QY 268 RLVYNGKPVFLKFGKIEEPVVGQGTFFYLMIKDFNLKLVIANNSFRSHYPSSEMD 327
DB 268 RLVYNGKPVFLKFGKIEEPVVGQGTFFYLMIKDFNLKLVIANNSFRSHYPSSEMD 327
QY 284 KFLINNKPPYFGFGHEDTPIINGRGNFASVMMFNLKMGANSFRFAHYPSSELMR 343
DB 284 KFLINNKPPYFGFGHEDTPIINGRGNFASVMMFNLKMGANSFRFAHYPSSELMR 343
QY 328 LADRLGLIVIDEAPHVIGITRYHNPETOKIAED-----EDNIRMRDR 368
DB 328 LADRLGLIVIDEAPHVIGITRYHNPETOKIAED-----EDNIRMRDR 368
QY 344 LADRLGLIVIDEAPHVIGITRYHNPETOKIAED-----EDNIRMRDR 400
DB 344 LADRLGLIVIDEAPHVIGITRYHNPETOKIAED-----EDNIRMRDR 400
QY 369 HKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAPDERTRV 427
DB 369 HKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAPDERTRV 427
QY 401 DKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAPDERTRV 458
DB 401 DKHPSVIMSVANBESNHPDABGFFKALYETANEMDRTRPVVMSMDAPDERTRV 458
QY 428 ALKYFDICVNRYYGWIYQGRIBEGLOALEKDIIEELYARHR-----KPIFTEFGADA 483
DB 428 ALKYFDICVNRYYGWIYQGRIBEGLOALEKDIIEELYARHR-----KPIFTEFGADA 483

```

```
Db 459 VAEILDVALNRKMYPDGDLLEAKVHLR---QEFHAMNRCGCKPIMITEYGADIVA 515
Qy 484 GIHYDPQMFSEBYOAEIVKTIIRLLKKDYIIGHVAFADFKTPQVNRPRILNHKGVF 543
Db 516 GFHDIDPVMFTBEEYVEYYQANHVVDFEFENFVGQANNFADFATSGVMRYOGNKKGVF 575
Qy 544 TRDRPKLVANVLRRLMSEV 563
Db 576 TRDRPKLAAHVFRERWTNI 595

RESULT 4
US-09-270-957-2
; Sequence 2, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270.957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRF
; ORGANISM: Bacillus sp.
US-09-270-957-2

Query Match 31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.2e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy 1 MYRPNRKKRFFILNGVWNLV-----TSKDRPIAVPGSMN--QYODLC 44
Db 1 MLYPINTETRGVFDLNGVWNLFDYKGLEEKWYESKLTDTISMVPSYNDIGVTKER 60
Qy 45 YEEGPFYKTFYVYPKLSQKHIRLYFAAVNTDCEVFLNGEKYGENHIEYLPEFVDYTGK 104
Db 61 NHIGVWYEREFYVAYLKDQRIYVRFSGATHKAIYVYNGELVYHKGGLPFPEALINNS 120
Qy 105 VKSGENELRVVVENLKYGFPFSKVPDSGTHTVGFGS-----PPANPDFPYG 154
Db 121 LRDGNRRVTVAVDNL-----DSTLPVGLYSERHEGLKVIRNKNPDPFFVYA 170
Qy 155 GIIRPVLEFTHARILDIWDTSESEPEKKLGKVKVIESEAVGQEMTKLG--FEE 212
Db 171 GIHRPVKLYTTPFYVEDISVTDNNGP--TGVTYTVDFQ---GKAETVKSVDDE 223
Qy 213 KIRTSNRFVGEFELLENARFWSLEDPLYLPLKVELKDEYTLDI-----GIRTSWDEK 267
Db 224 GKVASTEGLSGNVEIPVYILMEPLNTYLYQIKVELVNDGLTIDYEEBPGVRYTEVNDG 283
Qy 268 RLYLNGKVPFLKGFQKHEEPVYLGQGTFRPLMIKDPNLKMINANSFRTSHYSEEMLD 327
Db 284 KFLINNKPFYFGFKHEDTPIINGRGFNEASVNDMFLIKMIGANSFRTAHPYSEELMR 343
Qy 328 LADRLGILVIDEAPVIGTIRHYN-----PETOKIA-----ENIRMRIDR 368
Db 344 LADREGLVVIDETPAVGV---HLNFMATTGLGEGSERVSTWEKIRTFEHHQDVLELVS 400
Qy 369 HKNHPSVIMSVANPESENHPDAEGFKALEYETANEMD--RTRPVVMSMMDAPDERTDV 427
Db 401 DKNHPSVVMMSIANDAEATEEGAYEYFKPLVELTELDPQKRPTVILVFMAWTPR--TDK 458
Qy 428 ALKTYDVCNRRYGYWYIYQGRIEBGLQALEKDIEELVARRH---KPIPTYEGADALIA 483
Db 459 VAEILDVALNRKMYPDGDLLEAKVHLR---QEFHAMNRCGCKPIMITEYGADIVA 515
Qy 484 GIHYDPQMFSEBYOAEIVKTIIRLLKKDYIIGHVAFADFKTPQVNRPRILNHKGVF 543
```

```
Db 516 GFHDIDPVMFTBEEYVEYYQANHVVDFEFENFVGQANNFADFATSGVMRYOGNKKGVF 575
Qy 544 TRDRPKLVANVLRRLMSEV 563
Db 576 TRDRPKLAAHVFRERWTNI 595

RESULT 5
US-09-270-957-8
; Sequence 8, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270.957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 602
; TYPE: PRF
; ORGANISM: Bacillus sp.
US-09-270-957-8

Query Match 31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.2e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy 1 MYRPNRKKRFFILNGVWNLV-----TSKDRPIAVPGSMN--QYODLC 44
Db 1 MLYPINTETRGVFDLNGVWNLFDYKGLEEKWYESKLTDTISMVPSYNDIGVTKER 60
Qy 45 YEEGPFYKTFYVYPKLSQKHIRLYFAAVNTDCEVFLNGEKYGENHIEYLPEFVDYTGK 104
Db 61 NHIGVWYEREFYVAYLKDQRIYVRFSGATHKAIYVYNGELVYHKGGLPFPEALINNS 120
Qy 105 VKSGENELRVVVENLKYGFPFSKVPDSGTHTVGFGS-----PPANPDFPYG 154
Db 121 LRDGNRRVTVAVDNL-----DSTLPVGLYSERHEGLKVIRNKNPDPFFVYA 170
Qy 155 GIIRPVLEFTHARILDIWDTSESEPEKKLGKVKVIESEAVGQEMTKLG--FEE 212
Db 171 GIHRPVKLYTTPFYVEDISVTDNNGP--TGVTYTVDFQ---GKAETVKSVDDE 223
Qy 213 KIRTSNRFVGEFELLENARFWSLEDPLYLPLKVELKDEYTLDI-----GIRTSWDEK 267
Db 224 GKVASTEGLSGNVEIPVYILMEPLNTYLYQIKVELVNDGLTIDYEEBPGVRYTEVNDG 283
Qy 268 RLYLNGKVPFLKGFQKHEEPVYLGQGTFRPLMIKDPNLKMINANSFRTSHYSEEMLD 327
Db 284 KFLINNKPFYFGFKHEDTPIINGRGFNEASVNDMFLIKMIGANSFRTAHPYSEELMR 343
Qy 328 LADRLGILVIDEAPVIGTIRHYN-----PETOKIA-----ENIRMRIDR 368
Db 344 LADREGLVVIDETPAVGV---HLNFMATTGLGEGSERVSTWEKIRTFEHHQDVLELVS 400
Qy 369 HKNHPSVIMSVANPESENHPDAEGFKALEYETANEMD--RTRPVVMSMMDAPDERTDV 427
Db 401 DKNHPSVVMMSIANDAEATEEGAYEYFKPLVELTELDPQKRPTVILVFMAWTPR--TDK 458
Qy 428 ALKTYDVCNRRYGYWYIYQGRIEBGLQALEKDIEELVARRH---KPIPTYEGADALIA 483
Db 459 VAEILDVALNRKMYPDGDLLEAKVHLR---QEFHAMNRCGCKPIMITEYGADIVA 515
Qy 484 GIHYDPQMFSEBYOAEIVKTIIRLLKKDYIIGHVAFADFKTPQVNRPRILNHKGVF 543
Db 516 GFHDIDPVMFTBEEYVEYYQANHVVDFEFENFVGQANNFADFATSGVMRYOGNKKGVF 575
Qy 544 TRDRPKLVANVLRRLMSEV 563
Db 576 TRDRPKLAAHVFRERWTNI 595
```

RESULT 6
 US-09-270-957-15
 ; Sequence 15, Application US/09270957
 ; Patent No. 6641996
 ; GENERAL INFORMATION:
 ; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
 ; TITLE OF INVENTION: MICROBIAL-GLUCURONIDASE GENES, GENE
 ; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
 ; FILE REFERENCE: 190106.405C1
 ; CURRENT APPLICATION NUMBER: US/09/270.957
 ; EARLIER FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 112
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 15
 ; LENGTH: 602
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 US-09-270-957-15

Query Match 31.0%; Score 929; DB 4; Length 602;
 Best Local Similarity 35.6%; Pred. No. 2.2e-73;
 Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQNRKRFILINGVWNLV-----TSKDRPIAVPGSWE--QYQDLC 44
 DB 1 MLYPINTETRGVPDLNGVWNLV-----TSKDRPIAVPGSWE--QYQDLC 60
 QY 45 YEEGPFYKTFYVPPKLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDYTGK 104
 DB 61 NHIGVWYEREFVPAVLKQRIYVLRFGSATHKAIYVNGELVYEHKGGFLPFEALINNS 120
 QY 105 VKSGENELRVYVENRLKVGFSKVPDGGTHTVGFSG-----PPANDFPPYG 154
 DB 121 LRDKNRVTVAVADNLI-----DSTLPVGLYSEHREBGLKVIIRNKPDPFNVA 170
 QY 155 GIIRPVLIEFTDARILDIWDTSESEPEKLGKVKYKIEVEANQOEMTIKLG--EE 212
 DB 171 GLHRPVKIYTPPTTYVEDISVTDENG--TGYVTVYVDQ--GKAETKVSVDDE 223
 QY 213 KKIITSNRFEVGEFIIENARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTI 267
 DB 224 GKVAASLEGSGNVEIPVILMEPLNTYLYQIKVELVNDGLTIDVEEPPGVRTVEVNDG 283
 QY 268 RLYLNKRPVFLKFGKHEEPVVLGGCTFYPLMIKDFNLKWINANSFRTSHYPISEWLD 327
 DB 284 KFLINNKRPYFGFGKHEDTPIINGRGFNEASVMDFNILKWINANSFRTSHYPISELMR 343
 QY 328 LADRLGILVIDEAPHVIGITRYHVN-----PETOKIA-----EDNIRRMIDR 368
 DB 344 LADREGLVVIDETPAVG--HLMFMATTGLBGSERVSWEKIRTEHHODVLRVSR 400
 QY 369 HKHNPVIMSVANBESNHPDAEGFKALYETANEMD--RTRPVVWSMMDADPDETRDV 427
 DB 401 DKHNPVIMSVANBESNHPDAEGFKALYETANEMD--RTRPVVWSMMDADPDETRDV 458
 QY 428 ALKYPIDVCAVRVYGYIYQRIEGLQALEKDIIEELVYARH--KPIFTVEFGADATA 483
 DB 459 VAEILIDVIALNRKNGYFDGDLBAKVHLR--OEFHANMKRCGKPIIMITEYGADTVA 515
 QY 484 GIHYDPQWSESEYOAELEVEKTRILLLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVF 543
 DB 516 GFHIDIDVMTFEEYQVEYIQAANHVVFDENFVGEQAMNFAFPATISQGVNRVQGNKGVF 575
 QY 544 TRDRQPKLVAAVTLRLMSEV 563
 DB 576 TRDRKPKLAAHVFRERWTNI 595

RESULT 7
 US-09-149-727-4
 ; Sequence 4, Application US/09149727

Patent No. 6391547
 ; GENERAL INFORMATION:
 ; APPLICANT: Jefferson, Richard A.
 ; APPLICANT: Kilian, Andrzej
 ; APPLICANT: Keesee, Paul Konrad
 ; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
 ; TITLE OF INVENTION: USES THEREOF
 ; FILE REFERENCE: 190106.405
 ; CURRENT APPLICATION NUMBER: US/09/149.727
 ; EARLIER FILING DATE: 1998-09-08
 ; EARLIER APPLICATION NUMBER: US 60/058,263
 ; EARLIER FILING DATE: 1997-09-09
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 618
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 US-09-149-727-4

Query Match 31.0%; Score 929; DB 3; Length 618;
 Best Local Similarity 35.6%; Pred. No. 2.3e-73;
 Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQNRKRFILINGVWNLV-----TSKDRPIAVPGSWE--QYQDLC 44
 DB 17 MLYPINTETRGVPDLNGVWNLV-----TSKDRPIAVPGSWE--QYQDLC 76
 QY 45 YEEGPFYKTFYVPPKLSQKHRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDYTGK 104
 DB 77 NHIGVWYEREFVPAVLKQRIYVLRFGSATHKAIYVNGELVYEHKGGFLPFEALINNS 136
 QY 105 VKSGENELRVYVENRLKVGFSKVPDGGTHTVGFSG-----PPANDFPPYG 154
 DB 137 LRDKNRVTVAVADNLI-----DSTLPVGLYSEHREBGLKVIIRNKPDPFNVA 186
 QY 155 GIIRPVLIEFTDARILDIWDTSESEPEKLGKVKYKIEVEANQOEMTIKLG--EE 212
 DB 187 GLHRPVKIYTPPTTYVEDISVTDENG--TGYVTVYVDQ--GKAETKVSVDDE 239
 QY 213 KKIITSNRFEVGEFIIENARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTI 267
 DB 240 GKVAASLEGSGNVEIPVILMEPLNTYLYQIKVELVNDGLTIDVEEPPGVRTVEVNDG 299
 QY 268 RLYLNKRPVFLKFGKHEEPVVLGGCTFYPLMIKDFNLKWINANSFRTSHYPISEWLD 327
 DB 300 KFLINNKRPYFGFGKHEDTPIINGRGFNEASVMDFNILKWINANSFRTSHYPISELMR 359
 QY 328 LADRLGILVIDEAPHVIGITRYHVN-----PETOKIA-----EDNIRRMIDR 368
 DB 360 LADREGLVVIDETPAVG--HLMFMATTGLBGSERVSWEKIRTEHHODVLRVSR 416
 QY 369 HKHNPVIMSVANBESNHPDAEGFKALYETANEMD--RTRPVVWSMMDADPDETRDV 427
 DB 417 DKHNPVIMSVANBESNHPDAEGFKALYETANEMD--RTRPVVWSMMDADPDETRDV 474
 QY 428 ALKYPIDVCAVRVYGYIYQRIEGLQALEKDIIEELVYARH--KPIFTVEFGADATA 483
 DB 475 VAEILIDVIALNRKNGYFDGDLBAKVHLR--OEFHANMKRCGKPIIMITEYGADTVA 531
 QY 484 GIHYDPQWSESEYOAELEVEKTRILLLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVF 543
 DB 532 GFHIDIDVMTFEEYQVEYIQAANHVVFDENFVGEQAMNFAFPATISQGVNRVQGNKGVF 591
 QY 544 TRDRQPKLVAAVTLRLMSEV 563
 DB 592 TRDRKPKLAAHVFRERWTNI 611

RESULT 8
 US-09-149-727-8
 ; Sequence 8, Application US/09149727
 ; Patent No. 6391547

```

? GENERAL INFORMATION:
? APPLICANT: Jefferson, Richard A.
? APPLICANT: Kilian, Andrezej
? APPLICANT: Keese, Paul Konrad
? TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
? FILE OF INVENTION: US5 THEREOF
? FILE REFERENCE: 190106.405
? CURRENT APPLICATION NUMBER: US/09/149,727
? CURRENT FILING DATE: 1998-09-08
? EARLIER APPLICATION NUMBER: US 60/058,263
? EARLIER FILING DATE: 1997-09-09
? NUMBER OF SEQ ID NOS: 71
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 8
? LENGTH: 607
? TYPE: PRT
? ORGANISM: Bacillus sp.
? US-09-149-727-8

```

Query Match	30.8%	Score 925	DB 3	Length 607
Best Local Similarity	35.7%	Pred. No. 5.1e-73		
Matches 220	Conservative 92	Mismatches 223	Indels 82	Gaps 15

Qy	4	PQNKRRPILIIINGVNNLEV-----TSKORPIAVGCSNME--QYODCLCYEE	47
Db	9	PINTETRGVFDLNGVNNFKLDYDGKLEEMKYESKLTDTISMAVBSYNDIGVTEINNH1	68
Qy	48	GPRTYKTFPVYPKLSQKHIRLYFAFVAVNTDCEVFLNGEKEVENHIEYLPPFEDVTGAKYS	107
Db	69	GYYMYEREFTVAYLNDQRIVLRFSGATTHKALVYVNGSLVYEHKGGFLPFFAELNNSLRD	128
Qy	108	GEHELAVVYBNRLKVGCFPSKYPDGCTHTVGFPGS-----PPRANDFPPYGGIT	157
Db	129	GMRVTVAAVDNLI-----DDSTLPVGLYSERHEEGTGKYLKRNKPNDFPFNYAGLH	178
Qy	158	RPVLIEFTDHARLIDWDVTSESEPKKLGKVKVIEVSEBAVGOMTIKLG--EEBKXI	215
Db	179	RPKAITTTPTPYVEDSDVTPDNGP---TGIVTTYVDPO---GAAETKSVVDDEGKV	231
Qy	216	RTSNRFEVEGEFILLENARFWSLEDPYLPLPKVELEKDEYTLDI-----GIRTISWDEKRLY	270
Db	232	VASTEGLSGNVEIPNIVLWEPINTLYLTQGLKVELVNDGTLTIDYVEBPGRVVEVNDQKFL	291
Qy	271	LNGKPIFLGFGGHNHEFPVLGCGTGYPLMIKDPLIKITANNSTRISHYPRSEEMLDLAD	330
Db	292	INNKPEYFGFGGHEBTPINGRGFNEASVNDFNILIKIGANSFRTAHYPRSEELMLAD	351
Qy	331	RLGLIIVLIDRAPHNGIRTRYUN-----PEQOKLA-----EDNIRRMIDHKH	371
Db	352	REGVLVIDETPAVG---HLNPMATTTGEGESERVSTTEKIRTEBHODVLEIISHDKN	408
Qy	372	HPSVIMSWVAABESNHPDAEGFPKALYETANEMD--RTRPVVWSMDADPERTDRTVALK	430
Db	409	HPSVMMWSIANEATEEGBAGYEFKAPVELTKEIDPQKRPVTIYLVFWMAPE--TDKVAE	466
Qy	431	YPRIVCVNRVYQWYIYQGRIEEGDLAKEDKIDIELYARRK---KPIYTERGADALAGIH	486
Db	467	LIDVIALNNTNYQWYPDGGLLEAKVHLR---QEFANMKKCRPGKPIIMITEGADYVAGFH	523
Qy	487	YDPQWFSSEYOAELEVKIRLLKLDKDIYIGTHWAFADFPTPONVRRPILNKHGVTRD	546
Db	524	DIDPVWFTSEYQVEYVYQANHVVYDFEFENFVEBQAMNFADFATISQGVNRVQGNKKGVPTRD	583
Qy	547	ROPKLVAAHVLRLMSEV	563
Db	584	RKKKLAHVFRBRMTNI	600

RESULT 9
US-09-270-957-18
; Sequence 18, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:

: APPLICANT: Richard A. Jefferson and Jorge E. Mayer
 : TITLE OF INVENTION: MICROBIAL-GLUCONIDASE GENES, GENE
 : TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
 : FILE REFERENCE: 190106.405C1
 : CURRENT APPLICATION NUMBER: US/09/270,957
 : CURRENT FILING DATE: 1999-03-17
 : NUMBER OF SEQ. ID NOS: 112
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ. ID NO 18
 : LENGTH: 607
 : TYPE: PR1
 : ORGANISM: *Bacillus* sp.
 : US-09-270-957-18

Query Match	30.8%	Score 925;	DB 4;	Length 607;
Best Local Similarity	35.7%	Pred. No. 5.1e-73;		
Matches 220; Conservative	92;	Mismatches 223;	Indels 82;	Gaps 15

```

QY      4  PQRKKKFFIILINVMNMLEV-----TSKORPIAVPSSNME--QYODLCYBE 47
      9  PINTETGVEDPLNVMNMFKLDYGGKLEBKWYESKLDTTISNAVBSSYNDIGVTKTIRNHI 68
QY      48  GPFTYKTTFFVKKLSQKHRIYLFEAAVNTDCEVPLNGEKVGNHIEYLFEFVDVYGVKS 107
      69  GYVWYEREEFTVPAYLKQORIVLRGSAATHAKAIIVYNGSLVYEHKGFLPFEKIEINNSLD 128
QY      108  GENELRVVNEKRLKVGEPKRVPRDSGHTTVCFPGS-----FPFANDFPFYGGIT 157
      129  GMRKVTAAVNDIL-----DDSTLPVGLYSRHEBGLGKVIIRKNPNDFPFYAGI 178
QY      158  RPVLIEFTDARILIDIVDTSSEPEKKLGKVKYKIEVSEAVGOEMTICKG--BEKKI 215
      179  RPVKIYTTPTFYVDISIVTDFNPR--TGVTYTVDPQ----GKAETVKSVDVEBGV 231
QY      216  RTSNRFEVGEFIEENARFWSIEDYLYPLKYLEKDEYTLDI---GIRTSWDEKRLY 270
      232  VASTEGISGVNVEIPNVILMEPLNTLYLOIKVELVNDGLTIDVYEERPFVTRTEVNDGKFL 291
QY      271  LINGRPVLKPGKGKEEPVVGOGFFPYLMIDPFLKKNINNSPFTSHYVSEEMLDLAD 330
      292  INNKPEFFPKGEGKHEDTPINGRGNEASNVMDFNILKIGNSFTTAYIPSEELMRLLAD 351
QY      331  RLGLVLIDEAPHVGITREYHN-----PETOKIA-----EDNIRRMIDRHKN 371
      352  REGVLVIDEIPAVGV--HLNFMATITGEGSEKSVTWEKIRTEBHQDVLRELVSRDN 408
QY      372  HPSYIMMSVANPEPSNHPDAEGFKALYETANEND-RTRPVVMSAMDAERDERTVDALK 430
      409  HPSVYVMSIINAEATEBEGAYEYFKPLVELTKELDPQKRPVTIIVLFVATPE--TDKYAE 466
QY      431  YPDIVCNRRYYGWYIQGRIEBGLQALEKDIIEELYAHNR-----KPIFTEGADAIAGIH 486
      467  LIDIVIALNRINGMYFDGGDLEBAALVHLR--QESHANNKRCRPGKRMITEEGADLVVAGFH 523
QY      487  YDBPQWFSSEBYQALEVEKTRILRLKDUYIIGTHWAPADFEKTPONVRRLINHKGVFTRD 546
      524  DIDPVNTEEEYQVEYYQANHVVEDEFENFVGEOAMNPFADFATISQVNRVQCNKKGVFTRD 583
QY      547  RQKPLVAHVLRRLMSEV 563
      584  RKPEGLAHVFERERTNI 600

```

RESULT 10-
 US-09-270-957-28
 ; Sequence 28, Application US/09270957
 ; Patent No. 664196
 ; GENERAL INFORMATION:
 ; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
 ; TITLE OF INVENTION: MICROBIAL - GLUCONIDASE GENES, GENOMES
 ; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
 ; FILE REFERENCE: 190106.405C1
 ; CURRENT APPLICATION NUMBER: US/09/270,957

```

; RESULT 11
; US-09-862-660-2
; Sequence 2, Application US/09862660
; Patent No. 6664097
; GENERAL INFORMATION:
; APPLICANT: Russell, William
; APPLICANT: Klaenhammer, Todd
; TITLE OF INVENTION: LACTOBACILLUS BETA-GLUCURONIDASE AND DNA ENCODING THE SAME
; FILE REFERENCE: 5051.514
; CURRENT APPLICATION NUMBER: US/09/862,660
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/206,372
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.0

```

```

RESULT 12
US-09-715-858-4
; Sequence 4, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; TREATMENT OF LYSOSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/09/715,858
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 648

```

TYPE: PRT
ORGANISM: Mus musculus
US-09-715-858-4

Query Match 30.0%; Score 899.5; DB 4; Length 648;
Best Local Similarity 35.9%; Pred. No. 1e-70;
Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;

QY 1 MVRQRNKKRPIILLINGVWNLV-TSKDR-----PIAVGSNNQ 39
DB 27 MLFKESSSRRLKLDGLMFRADLSNNRLOQFECQWROPRESGPVLDMPVPSFNDI 86
QY 40 YQDLCYEE--GPFYKTFEYVPKLSQ--KHIRLYFAAVNTDCEVFLNGEKVGNHLEY 94
DB 87 TQBALRDFIGWYERAILPRRTQDTDMRVLRINSAYAVVWNGHVVHEGCH 146
QY 95 LPEFVDVTKVYSG--ENELRVVENRLKVGGPSPKVP-----DSGTHVGFPGSF 143
DB 147 LPEFADISKLVQSGFLTTCRITAIINNTLT---PHTLPGRIVYKTDTSVMPKGYF--V 200
QY 144 PRANPDPFPGGIRPVLIEFTDARILDIWDTSESPEKKLGKVK--VKIEVSEAVG 201
DB 201 QDTSEDFPNVAGLHRSVVLVYTPPTTYIDITVITN---VEODIGLVTYMISVQSEHF-- 255
QY 202 QEMTIKLGEEBKRTSNRPFVEGEFLENNARFW---SLEDP-VLYPLKV-----ELEKD 251
DB 256 -QLEVOQLDEGSKVVAHGTGNOQLQVPSANLWMPYLMHEHPAYMSLEVKYTTTESVTD 314
QY 252 EYTLDIGIRITISWDEKRLYLNGKPVFLKFGKHEFPVLGGTFYPLMIKDFNLKMINA 311
DB 315 YTLPIGIRYAVVTKSKFLINGKPFYFGVANKHESDRIKSGFDMPLVVKDFNLRLWIGA 374
QY 312 NSFTSHYPRSEEMLDLDRGLVYIDEAHYVGT-----RYHYNETOKIAEDN 361
DB 375 NSFTSHYPRSEEMVQLCDRYGIWIDECPGVGIPLPOSFGNESLRHNL-----EVBEL 429
QY 362 IRRMIDRHKNPSYIMVSNAPESNHPDAGFFKALYETANEMDRTPVWVS--NMDA 419
DB 430 VR----RDKNHPAVWMSVSNAPSSALKPAAVYFETLITHTKALDLTRPVFVSACKYDA 485
QY 420 PDERTDVALKYPDIVCNARYYGYWYIGRIEGLQALEKDIIEELVYARHKPIFVTEFGA 479
DB 486 -----DLGARYVGVICNSYFWSYHDYGHLEVIQPOLNSQFEMMYKTHQPIIOSEYGA 539
QY 480 DALGCIHYDPQMSSEYQAELEVEKTRLL--LKKDYIIGHVAFADFKTPQNVRRPIL 537
DB 540 DALGCIHEDPRPMSSEYQAKVLENYHSLVDQKREYVVGELIWFADFMTNQSPLVYIG 599
QY 538 NHKGVPTDRQPKLVAYHLR-RLW 560
DB 600 NKGIFTRQROPKTSAPILRERYW 623

RESULT 13
US-09-118-276-12
Sequence 12, Application US/09118276
Patent No. 6693185

GENERAL INFORMATION:
APPLICANT: BABYCHUK, ELENA;
APPLICANT: KUSHNIR, SERGEI;
APPLICANT: DE BNOIR, MARC;
APPLICANT: INZE, DIRK
TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIXBEY, FRIEDMAN, LEEBOM, & FERGUSON
STREET: 8180 GREENSBORO DRIVE, SUITE 800
CITY: MCLEAN,
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22102
COMPUTER READABLE FORM:

MEDIUM TYPE: 3-1/2" DISKETTE
COMPUTER: IBM-COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/118,276
FILING DATE: 17-JUL-1998
PRIOR APPLICATION DATA: NONE
ATTORNEY/AGENT INFORMATION:
NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
REGISTRATION NUMBER: 31,196; 43,077
REFERENCE/DOCKET NUMBER: 6201-0003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 790-9110
TELEFAX: (703) 883-0370
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1010 RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
US-09-118-276-12

Query Match 29.4%; Score 882; DB 4; Length 1010;
Best Local Similarity 34.0%; Pred. No. 6.9e-69;
Matches 206; Conservative 109; Mismatches 223; Indels 62; Gaps 11;

QY 1 MVRQRNKKRPIILLINGVWNLV-----TSKDRPIAVGSNNQYQDLCY 45
DB 408 MVRPFETFTREIKKLDGLMFRADLSRENGCIDPRWMSALQSRALIAVSGSFNDQPADADI 467
QY 46 EE--GPFYKTFEYVPKLSQKHIRLYFAAVNTDCEVFLNGEKVGNHLEYLPEFVDYTG 103
DB 468 RNYAGNVWYQREVEPIPKMGAGQRIVLREDAVTHYKGVWVNNQVEHOGGYTPFEADVTP 527
QY 104 KYKSGEN-ELRVVENRLKVGGPSPKV---PDSGTHVGFPGSFPPANFDFPYGGRIRP 159
DB 528 YVINGKSRITVVCNNELNMOTIPPGWYITDENGKKKOSYF-----HDFPNYAGIHS 580
QY 160 VLIBFTDARILDIWDTSESPEKKLGKVKVKEVSEAVGOEMTIKLGEEBKRTSN 219
DB 581 VMLYTPPTWVDITVYVTHVAQ---DCNHSVDMQV--VANGDVSVELRADQGVANG 634
QY 220 RVEGEFLENNARWLSLEDPLVLPV---KYLEKDEYTLDIGIRITISWDEKRLYLNGK 275
DB 635 QGTSGLQVWVPHLMQPEBGLYLECYAKSQTECDIYPLRVGIRSVAVKGEOPLINHKP 694
QY 276 VFLKPGKHEFPVLGGTFYPLMIKDFNLKMINANSFRTSHYPSSEEMLDLDRGLIL 335
DB 695 FYTFGFGHEDADLRGKGFNDVLMVHDHALMDWIGANSYRTSHYPVABEKLMDWADDEHGI 754
QY 336 VIDEAPHYGI-----TRYHYNETOKIAEDNIRRMIDRHKNPSYIM 377
DB 755 VIDEFAVGFNLISGIFGEAGNKPKEYLSEBAVNGEFOALHQAIMEKLIADKHPVYM 814
QY 378 WSVANEPESNHPDLBGFPPKALYETANEMDRTPVWVSMM--DAPDETRDVALKYPDIV 435
DB 815 WSIANEPDTPQGRAREFYAPLAETRKLDPFRPITCVVMMCDATDTDISL---FDVL 870
QY 436 CVNRYGYVYIGRIEGLQALEKDIIEELVYARHKPIFVTEFGDALGCIHYDPQMSSE 495
DB 871 CLNRYGYVQSGLETAKEVLEKELAMQEKLHQPITTEYGVDTLAGLSMYTDMWSE 930
QY 496 EYQAELEVEKTRILLKKDYIIGHVAFADFKTPQNVRRPILNHKGVPTDRQPKLVAYH 555
DB 931 EYQCAWLDMYHRVDRSAVAVGEVWNPADPATSGILRVGNNKKGIFTRRKRKSAFL 990
QY 556 LRLRLS 561
DB 991 LQKRW 996

RESULT 14

```

US-09-149-727-6
; Sequence 6, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kiliian, Andrej
; APPLICANT: Keesse, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 603
; TYPE: PR1
; ORGANISM: Escherichia coli
US-09-149-727-6

```

```

Query Match      29.3%; Score 879; DB 3; Length 603;
Best Local Similarity 33.8%; Pred. No. 5,8e-69;
Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

```

```

QY 1 MVRPQRNKKRFFILLNGVNNLEV-----TSKDRPIAVPGSNMEQYDLCY 45
DB 1 MVRPVTPTREIKKLDGLMAFSLDRNCGIDQRMWESALQESRALAVPGSFNDQFADADI 60
QY 46 EE--GPFYKTFVVPKXLSOKHRLYFAAVNTDCEVFLNGKVGKGNHIEYLPEVDYTG 103
DB 61 RNVAQNWYQREVIFPKMGAGQRIYLRFDVAVTHYGKAVNNQEWMEHOGGTYRFEADVTP 120
QY 104 KYKSGEN-ELRVVENRLKVGGFPSKY---PDSGTHTVGFGSPFPANDFPPYGGIIRP 159
DB 121 YVLAGKSVRLTVCCNNELMWQITPRGMVITDENGKKQSYF-----HDFVYAGIHR 173
QY 160 VLIEFTDARLIDIVDTSESEPEKKLGKVKYKIEVSEAVQGMITKIGEEKKIRTSN 219
DB 174 VMLYTPNTWVDITVTVTHVAQ---DCNHAQVDMQV---VANGDVSVELRDADQOVVATG 227
QY 220 RFVGEFTELENARFMSLEDPYLYPL-----KVELEKDEYTLIDIGRTISWDEKRLYLNGKP 275
DB 228 QGTSGLQVNVPHLMQPGEGYLYELCVTAKSQTECDIYLRVGRSAVAVGEOFLINHKP 287
QY 276 VFLKGFGEKHEEPVLAGQTFYPLMIKDFNLKMINANSFRITSHYPYSEEWLADRLGIL 335
DB 288 FYFTGFGHEDADLRGKGFQNVLMVHDALMDWIGANSYRTSHYPYAEEMLDWADERGIV 347
QY 336 VIDEAHVGI-----TRYHNPEFQKLAEDNIRKMTDRHKNHPSVIM 377
DB 348 VIDETAAGVFNLSLIGFEAGNKKPELYSEAVNGEFOQAHLOAIKELIADKNHPSVIM 407
QY 378 MSVANEPSNHPDAEGFFKALYETANEMDRTRPVVWVSM--DAPDERTDVALKYFDIV 435
DB 408 WSIANEEDTRPGQAREYFAPLAETARKLDPTRITICVNVWFCAHNDTISL-----FDVL 463
QY 436 CVNRYYGWYIYQRIIEBGLQALEKDIIELYARHKRPFVTEFGADALAGIHYPQWSE 495
DB 464 CLNRYGMYVQSGDLETAKEVLEKELAMQEKHQPIITIEYGVDTLAGHSMYTDWSE 523
QY 496 EYQAEVLEKTRILLKKDYIIGTHVMAFADPTQNVNRRPILNHKGVTTRROKLVAVH 555
DB 524 EYQCAMLDMYHRVDRSAVAVGEOVMNFPADATSGILRVGKNGKGIIFTRDKRKSAAFL 583
QY 556 LRRILMS 561
DB 584 LQKRWT 589

```

RESULT 15
US-09-270-957-17

```

; Sequence 17, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; APPLICANT: Kiliian, Andrej
; APPLICANT: Keesse, Paul Konrad
; TITLE OF INVENTION: MICROBIAL-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 603
; TYPE: PR1
; ORGANISM: Escherichia coli
US-09-270-957-17

```

```

Query Match      29.3%; Score 879; DB 4; Length 603;
Best Local Similarity 33.8%; Pred. No. 5,8e-69;
Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

```

```

QY 1 MVRPQRNKKRFFILLNGVNNLEV-----TSKDRPIAVPGSNMEQYDLCY 45
DB 1 MVRPVTPTREIKKLDGLMAFSLDRNCGIDQRMWESALQESRALAVPGSFNDQFADADI 60
QY 46 EE--GPFYKTFVVPKXLSOKHRLYFAAVNTDCEVFLNGKVGKGNHIEYLPEVDYTG 103
DB 61 RNVAQNWYQREVIFPKMGAGQRIYLRFDVAVTHYGKAVNNQEWMEHOGGTYRFEADVTP 120
QY 104 KYKSGEN-ELRVVENRLKVGGFPSKY---PDSGTHTVGFGSPFPANDFPPYGGIIRP 159
DB 121 YVLAGKSVRLTVCCNNELMWQITPRGMVITDENGKKQSYF-----HDFVYAGIHR 173
QY 160 VLIEFTDARLIDIVDTSESEPEKKLGKVKYKIEVSEAVQGMITKIGEEKKIRTSN 219
DB 174 VMLYTPNTWVDITVTVTHVAQ---DCNHAQVDMQV---VANGDVSVELRDADQOVVATG 227
QY 220 RFVGEFTELENARFMSLEDPYLYPL-----KVELEKDEYTLIDIGRTISWDEKRLYLNGKP 275
DB 228 QGTSGLQVNVPHLMQPGEGYLYELCVTAKSQTECDIYLRVGRSAVAVGEOFLINHKP 287
QY 276 VFLKGFGEKHEEPVLAGQTFYPLMIKDFNLKMINANSFRITSHYPYSEEWLADRLGIL 335
DB 288 FYFTGFGHEDADLRGKGFQNVLMVHDALMDWIGANSYRTSHYPYAEEMLDWADERGIV 347
QY 336 VIDEAHVGI-----TRYHNPEFQKLAEDNIRKMTDRHKNHPSVIM 377
DB 348 VIDETAAGVFNLSLIGFEAGNKKPELYSEAVNGEFOQAHLOAIKELIADKNHPSVIM 407
QY 378 MSVANEPSNHPDAEGFFKALYETANEMDRTRPVVWVSM--DAPDERTDVALKYFDIV 435
DB 408 WSIANEEDTRPGQAREYFAPLAETARKLDPTRITICVNVWFCAHNDTISL-----FDVL 463
QY 436 CVNRYYGWYIYQRIIEBGLQALEKDIIELYARHKRPFVTEFGADALAGIHYPQWSE 495
DB 464 CLNRYGMYVQSGDLETAKEVLEKELAMQEKHQPIITIEYGVDTLAGHSMYTDWSE 523
QY 496 EYQAEVLEKTRILLKKDYIIGTHVMAFADPTQNVNRRPILNHKGVTTRROKLVAVH 555
DB 524 EYQCAMLDMYHRVDRSAVAVGEOVMNFPADATSGILRVGKNGKGIIFTRDKRKSAAFL 583
QY 556 LRRILMS 561
DB 584 LQKRWT 589

```

Search completed: January 24, 2005, 07:54:02
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 07:49:26 ; Search time 81 Seconds
(without alignments)
2511.184 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001
Sequence: 1 MVRPQRNKKRIRLLINGVNN.....TRDQPKLVAVHVRIMSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10D_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	14 US-10-364-649-6	Sequence 6, Appli
2	2999	99.9	563	14 US-10-364-649-21	Sequence 21, Appli
3	2999	99.9	563	14 US-10-369-493-2995	Sequence 2995, Ap
4	929	31.0	602	14 US-10-120-145-2	Sequence 2, Appli
5	929	31.0	602	14 US-10-364-649-2	Sequence 2, Appli
6	929	31.0	602	14 US-10-364-649-8	Sequence 8, Appli
7	929	31.0	602	14 US-10-364-649-15	Sequence 15, Appli
8	929	31.0	618	14 US-10-120-145-4	Sequence 4, Appli
9	925	30.8	607	14 US-10-120-145-8	Sequence 8, Appli
10	925	30.8	607	14 US-10-364-649-18	Sequence 18, Appli
11	925	30.8	615	14 US-10-364-649-28	Sequence 28, Appli
12	919.5	30.6	598	10 US-09-862-660-2	Sequence 2, Appli
13	919.5	30.6	598	15 US-10-673-935-2	Sequence 2, Appli

14	899.5	30.0	648	14 US-10-421-175-4	Sequence 4, Appli
15	882	29.4	1010	9 US-09-118-276-12	Sequence 12, Appli
16	882	29.4	1010	16 US-10-705-197A-12	Sequence 12, Appli
17	879	29.3	603	14 US-10-161-403-106	Sequence 106, App
18	879	29.3	603	14 US-10-120-145-6	Sequence 6, Appli
19	879	29.3	603	14 US-10-364-649-17	Sequence 17, Appli
20	879	29.3	603	14 US-10-364-649-23	Sequence 23, Appli
21	879	29.3	603	14 US-10-369-493-842	Sequence 842, App
22	879	29.3	603	17 US-10-161-408-17	Sequence 17, Appli
23	877	29.2	618	14 US-10-356-088-28	Sequence 28, Appli
24	877	29.2	618	16 US-10-799-326-28	Sequence 28, Appli
25	875	29.2	604	10 US-09-893-525-37	Sequence 37, Appli
26	875	29.2	604	17 US-10-763-180-37	Sequence 37, Appli
27	875	29.2	659	10 US-09-893-525-40	Sequence 40, Appli
28	875	29.2	659	17 US-10-763-180-40	Sequence 40, Appli
29	875	29.2	850	10 US-09-893-525-42	Sequence 42, Appli
30	875	29.2	850	17 US-10-763-180-42	Sequence 42, Appli
31	872.5	29.1	602	13 US-10-195-158-5	Sequence 5, Appli
32	872.5	29.1	602	14 US-10-195-158-5	Sequence 5, Appli
33	872.5	29.1	607	14 US-10-369-493-15291	Sequence 15291, A
34	872.5	29.1	711	14 US-10-338-411-23	Sequence 23, Appli
35	872.5	29.1	711	15 US-10-389-640-23	Sequence 23, Appli
36	865.5	28.8	613	14 US-10-120-145-5	Sequence 5, Appli
37	865.5	28.8	613	14 US-10-364-649-16	Sequence 16, Appli
38	865.5	28.8	651	14 US-10-421-175-2	Sequence 2, Appli
39	865.5	28.8	722	14 US-10-136-841-6	Sequence 6, Appli
40	865.5	28.8	722	15 US-10-272-531A-6	Sequence 6, Appli
41	865.5	28.8	722	15 US-10-272-483A-6	Sequence 6, Appli
42	705.5	23.5	450	14 US-10-356-088-40	Sequence 40, Appli
43	705.5	23.5	450	16 US-10-799-326-40	Sequence 40, Appli
44	692.5	23.1	376	14 US-10-364-649-4	Sequence 4, Appli
45	692.5	23.1	376	14 US-10-364-649-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1
US-10-364-649-6
; Sequence 6, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-364-649-6
Query Match 99.9%; Score 2999; DB 14; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.8e-241;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVRPQRNKKRIRLLINGVNNLEVTSSKDRPIAVPGSNMEQYODLCYEESGPFYKTTFTYVPK 60
DB 1 MVRPQRNKKRIRLLINGVNNLEVTSSKDRPIAVPGSNMEQYODLCYEESGPFYKTTFTYVPK 60
QY 61 XLGQKHRIPLVAAVNTDCEVFLNGEKVGENHIEVLPEVDVTGKVKSGENLRVAVENRL 120

```

Db      61  XLSQGHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKXKSGENSELRVVENRL 120
Qy      121  KVGFPSPKVPDSGHTHTVGFPGSPFPANFDFFPYGGIIRPVLIETDTHARILDIWDTSES 180
Db      121  KVGFPSPKVPDSGHTHTVGFPGSPFPANFDFFPYGGIIRPVLIETDTHARILDIWDTSES 180
Qy      181  EPEKLGKVKVKEVSEEAQVQEMTIKLGEEKKIRTSNRFVGEFIIENARFWSLEDPY 240
Db      181  EPEKLGKVKVKEVSEEAQVQEMTIKLGEEKKIRTSNRFVGEFIIENARFWSLEDPY 240
Qy      241  LYPFLKVELEKDEYTLIDIGIRTIISWDEKRLYLNGKVPFLKGFGEHPVLGOGTFYPLMT 300
Db      241  LYPFLKVELEKDEYTLIDIGIRTIISWDEKRLYLNGKVPFLKGFGEHPVLGOGTFYPLMT 300
Qy      301  KDFNLLKMIANSPFRTSHYPYSEEWLADRLGILVIDEAPHVIGIRHYHNPETQKIAED 360
Db      301  KDFNLLKMIANSPFRTSHYPYSEEWLADRLGILVIDEAPHVIGIRHYHNPETQKIAED 360
Qy      361  NIRRMIDRHKNHPSVIMSVANEPESNHPDABGFPALEYTANEMDRTPVVMVSMMDAP 420
Db      361  NIRRMIDRHKNHPSVIMSVANEPESNHPDABGFPALEYTANEMDRTPVVMVSMMDAP 420
Qy      421  DERTRDVALKYFDIVCVNRYYGWYIYQGRIEBGLQALEKDIIEELYARHRKPIFTEFGAD 480
Db      421  DERTRDVALKYFDIVCVNRYYGWYIYQGRIEBGLQALEKDIIEELYARHRKPIFTEFGAD 480
Qy      481  AIAGIHYDPPQMFSEEEYQAEIVERTIRLLKKDYIIGTHVMAFADFKTPQNVRRPILNKH 540
Db      481  AIAGIHYDPPQMFSEEEYQAEIVERTIRLLKKDYIIGTHVMAFADFKTPQNVRRPILNKH 540
Qy      541  GVFTDRQPKLVAVHVRRLMSEV 563
Db      541  GVFTDRQPKLVAVHVRRLMSEV 563

RESULT 2
US-10-364-649-21
; Sequence 21, Application US/10364649
; Publication No. US2003022921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCUCONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1) --(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-364-649-21

Query Match          99.9%; Score 2999; DB 14; Length 563;
Best Local Similarity 100.0%; Pred. No. 1,8e-241;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      61  XLSQGHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKXKSGENSELRVVENRL 120
Qy      121  KVGFPSPKVPDSGHTHTVGFPGSPFPANFDFFPYGGIIRPVLIETDTHARILDIWDTSES 180
Db      121  KVGFPSPKVPDSGHTHTVGFPGSPFPANFDFFPYGGIIRPVLIETDTHARILDIWDTSES 180
Qy      181  EPEKLGKVKVKEVSEEAQVQEMTIKLGEEKKIRTSNRFVGEFIIENARFWSLEDPY 240
Db      181  EPEKLGKVKVKEVSEEAQVQEMTIKLGEEKKIRTSNRFVGEFIIENARFWSLEDPY 240
Qy      241  LYPFLKVELEKDEYTLIDIGIRTIISWDEKRLYLNGKVPFLKGFGEHPVLGOGTFYPLMT 300
Db      241  LYPFLKVELEKDEYTLIDIGIRTIISWDEKRLYLNGKVPFLKGFGEHPVLGOGTFYPLMT 300
Qy      301  KDFNLLKMIANSPFRTSHYPYSEEWLADRLGILVIDEAPHVIGIRHYHNPETQKIAED 360
Db      301  KDFNLLKMIANSPFRTSHYPYSEEWLADRLGILVIDEAPHVIGIRHYHNPETQKIAED 360
Qy      361  NIRRMIDRHKNHPSVIMSVANEPESNHPDABGFPALEYTANEMDRTPVVMVSMMDAP 420
Db      361  NIRRMIDRHKNHPSVIMSVANEPESNHPDABGFPALEYTANEMDRTPVVMVSMMDAP 420
Qy      421  DERTRDVALKYFDIVCVNRYYGWYIYQGRIEBGLQALEKDIIEELYARHRKPIFTEFGAD 480
Db      421  DERTRDVALKYFDIVCVNRYYGWYIYQGRIEBGLQALEKDIIEELYARHRKPIFTEFGAD 480
Qy      481  AIAGIHYDPPQMFSEEEYQAEIVERTIRLLKKDYIIGTHVMAFADFKTPQNVRRPILNKH 540
Db      481  AIAGIHYDPPQMFSEEEYQAEIVERTIRLLKKDYIIGTHVMAFADFKTPQNVRRPILNKH 540
Qy      541  GVFTDRQPKLVAVHVRRLMSEV 563
Db      541  GVFTDRQPKLVAVHVRRLMSEV 563

RESULT 3
US-10-369-493-2295
; Sequence 2995, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2995
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-10-369-493-2295

Query Match          99.9%; Score 2999; DB 14; Length 563;
Best Local Similarity 99.8%; Pred. No. 1,8e-241;
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1  MVRPQRNKKRFFILLNGVWNLVETSKDRPIAVPGSWNEQYODLCYEBGPFTYKTFYVPK 60
Db      1  MVRPQRNKKRFFILLNGVWNLVETSKDRPIAVPGSWNEQYODLCYEBGPFTYKTFYVPK 60
Qy      61  XLSQGHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKXKSGENSELRVVENRL 120
Db      61  XLSQGHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKXKSGENSELRVVENRL 120

```

```

Qy      1  MVRPQRNKKRFFILLNGVWNLVETSKDRPIAVPGSWNEQYODLCYEBGPFTYKTFYVPK 60
Db      1  MVRPQRNKKRFFILLNGVWNLVETSKDRPIAVPGSWNEQYODLCYEBGPFTYKTFYVPK 60
Qy      61  XLSQGHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKXKSGENSELRVVENRL 120
Db      61  XLSQGHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKXKSGENSELRVVENRL 120
Qy      121  KVGFPSPKVPDSGHTHTVGFPGSPFPANFDFFPYGGIIRPVLIETDTHARILDIWDTSES 180
Db      121  KVGFPSPKVPDSGHTHTVGFPGSPFPANFDFFPYGGIIRPVLIETDTHARILDIWDTSES 180

```

Db 121 KVGFPSPKVPDSGTHVTFGSGFPANFDFPYGIIIRPVLEFTDHAIRLIDWDTSES 180
Qy 181 EPEKLGKVKIKIYSEEAAGQEMTIKLGEEBKIRTSNRPVEGFIENARFASLEDPY 240
Db 181 EPEKLGKVKIKIYSEEAAGQEMTIKLGEEBKIRTSNRPVEGFIENARFASLEDPY 240
Qy 241 LYPKLEKEDEYTLIDIGRTISWDEKRLYLNGKRVFLKFGKEEFPVLGGCTYPLMI 300
Db 241 LYPKLEKEDEYTLIDIGRTISWDEKRLYLNGKRVFLKFGKEEFPVLGGCTYPLMI 300
Qy 301 KDFNLKMINANSFRTSHYPSSEEWLADLRGLIVIDEAPHVIGITRHYNPETOKIAD 360
Db 301 KDFNLKMINANSFRTSHYPSSEEWLADLRGLIVIDEAPHVIGITRHYNPETOKIAD 360
Qy 361 NIRMIDHKKHPSIYMSVANEPESNPDAGFFKALYETANENDRTPVVMASMDAP 420
Db 361 NIRMIDHKKHPSIYMSVANEPESNPDAGFFKALYETANENDRTPVVMASMDAP 420
Qy 421 DERTRDVALKCFDIYCVARYGWYIYQGRIEEGLQALEKDIIELYARHRKPIFYTEFGAD 480
Db 421 DERTRDVALKCFDIYCVARYGWYIYQGRIEEGLQALEKDIIELYARHRKPIFYTEFGAD 480
Qy 481 AIAGIHYPDPOMFSEYQAELEVEKTIIRLLKKDYIIGTHVAFADFKTPQNVRRPILNHK 540
Db 481 AIAGIHYPDPOMFSEYQAELEVEKTIIRLLKKDYIIGTHVAFADFKTPQNVRRPILNHK 540
Qy 541 GVFTDRDQPKLVAVHVRRLMSEV 563
Db 541 GVFTDRDQPKLVAVHVRRLMSEV 563

RESULT 4

US-10-120-145-2
; Sequence 2, Application US/10120145
; Publication No. US20030157684A1
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/10/120,145
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-120-145-2

Query Match 31.0%; Score 929; DB 14; Length 602;
Best Local Similarity 35.6%; Pred. No. 1.5e-68;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
Qy 1 MVRPQNRKKRFFILLINGVWNLV-----TSKORPIVPSGSMN--QYDLC 44
Db 1 MLYPNTETRGVFDLNGVWNLFLDYKGLBEKWKESKLDITISMAVPSYNDIGVTKIR 60
Qy 45 YEEGFFYKTFYVYPKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHLEYLFEVDVYTGK 104
Db 61 NHIGVWYEREFYVAYLKQRIYVRFSGATHKAIIVYNGELVYHKGGLPFELEINNS 120
Qy 105 VKSGENELRVVVENLKVKGFPSPKVPDSGTHVTFGSG-----PPANPDPFPG 154
Db 121 LRDGNRVTVAVDNLT-----DDSTLPVGLYSRHEEGLKVIIRKNKNPDPFFNYA 170
Qy 155 GIIRPVLEFTDHAIRLIDWDTSESPEKLGKVKIKIYSEEAAGQEMTIKLG--BEE 212

Db 171 GLHRPVKITYTTFYVVEDISVTDENGP---TGIVTYVDPQ---GKAETVKASVVDSE 223
Qy 213 KKIRTSNRFVEGEFIENARFWSLEDPYLKPVLEKDEYTLDI-----GRTISWDEK 267
Db 224 GKVAVSTEGLSGNVEIPVUILMEPLNTLYQIKVELVNDGDLTIDYEEPPFGVRTAEVNDG 283
Qy 268 RLXYNGKRVFLKFGKEEFPVLGGCTYPLMIDFNLKMINANSFRTSHYPSSEWLD 327
Db 268 KFLNNKPFYFGFGKHEDTPTINGRGENASVNDPILKMINAASFRTAHYPSSEELMR 343
Qy 328 LADRGLIVIDEAPHVIGITRHYN-----PETOKIA-----EDNIRMIDR 368
Db 344 LADRGLIVIDETRAVGV---HINMATTTGLEGSESRVTEKIRTFEHHQDVLELYSR 400
Qy 369 HKNHSVIMSVANEPESNPDAGFFKALYETANEND--RTRPVVMSMDAPDERTDV 427
Db 401 DKNHPSVVMMSIANEAAETEEGAYEYFKPLVELTGLDLPQKRPPVITVLFVMAFPE--TDK 458
Qy 428 ALKTYDIYCVARYGWYIYQGRIEEGLQALEKDIIELYARHR---KPIFYTEFGADALA 483
Db 459 VAEILDVIALNRNGWYFDGDLLEAKVHLR---QEFHAMNKRCPGKPIMTTEYGADTVYA 515
Qy 484 GIHYDPDPOMFSEYQAELEVEKTIIRLLKKDYIIGTHVAFADFKTPQNVRRPILNHKGVF 543
Db 516 GFHDIDPMTFBEIYVEYTYQANHVVFDPEFNPFVGEQANFPADFATISQGVMYQSGKKGVF 575
Qy 544 TRDRQPKLVAVHVRRLMSEV 563
Db 576 TRDRKPKLAHVFRERMTNI 595

RESULT 5

US-10-364-649-2
; Sequence 2, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-2

Query Match 31.0%; Score 929; DB 14; Length 602;
Best Local Similarity 35.6%; Pred. No. 1.5e-68;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
Qy 1 MVRPQNRKKRFFILLINGVWNLV-----TSKORPIVPSGSMN--QYDLC 44
Db 1 MLYPNTETRGVFDLNGVWNLFLDYKGLBEKWKESKLDITISMAVPSYNDIGVTKIR 60
Qy 45 YEEGFFYKTFYVYPKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHLEYLFEVDVYTGK 104
Db 61 NHIGVWYEREFYVAYLKQRIYVRFSGATHKAIIVYNGELVYHKGGLPFELEINNS 120
Qy 105 VKSGENELRVVVENLKVKGFPSPKVPDSGTHVTFGSG-----PPANPDPFPG 154
Db 121 LRDGNRVTVAVDNLT-----DDSTLPVGLYSRHEEGLKVIIRKNKNPDPFFNYA 170
Qy 155 GIIRPVLEFTDHAIRLIDWDTSESPEKLGKVKIKIYSEEAAGQEMTIKLG--BEE 212

Db 171 GLHRPVKIYTPPTVEEDISVTDENG---TGTVTYVDFQ----GKAETVKSVDDE 223
 Qy 213 KKIIRTSNRFVEGEFLEENAFMSLEDPIYPLKVELEKDEYTLDI-----GIRTI5MDEK 267
 Db 224 GKVVASTEGISGNVEIPNVILMEPLNTLYXQIKVELVNDGLTIDVYEEBPGVRYEVNG 283
 Qy 268 RLYANGKPVFLKPGFKHEEPVLGGQTFYPLMKIDFNLLKWINANSFRSHYPSSEMD 327
 Db 284 KFLINNKPFYFGKGHEDTPINGRGFNEASVMDFNILKMGANSFRTHAYPSSELMR 343
 Qy 328 LADRLGIIVDEAPHVGTIRHYN-----PETOKIA-----EDNIRMIDR 368
 Db 344 LADREGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEXIKTFEHHQDVLRELVS 400
 Qy 369 HKHPSVIMSVANPEPSNHPDAEGFPKALYETANEMD--RTRPVVMVMDADERTRDV 427
 Db 401 DKHPSVIMSVANPEPSNHPDAEGFPKALYETANEMD--RTRPVVMVMDADERTRDV 458
 Qy 428 ALKYPDIVCNRYGYWYIYQRIEGLQALEKDIIEIYARH---KPIVTEFGADATA 483
 Db 459 VAEIIVIALNRNGWYFDGDLGAAKVHLR---QEFHANNKRCPCPKIMITEYADTVA 515
 Qy 484 GIHYDPQMFSEBYQALVEKTRIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGV 543
 Db 516 GFHDIDPVMTTEBYQVEYQANHVPDEFENFVGEQAMNPAFATSQGVWRVQGNKGV 575
 Qy 544 TRDRPKLVAAHVRRLMSEV 563
 Db 576 TRDRPKLVAAHVRERWTNI 595

RESULT 6
 US-10-364-649-8
 ; Sequence 8, Application US/10364649
 ; Publication No. US2003022921A1

GENERAL INFORMATION:

APPLICANT: Richard A. Jefferson and Jorge E. Mayer

TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE

FILE REFERENCE: 190106.405C1

CURRENT APPLICATION NUMBER: US/10/364,649

PRIOR FILING DATE: 2003-02-12

PRIOR APPLICATION NUMBER: 10/364,649

PRIOR FILING DATE: 2003-02-12

PRIOR APPLICATION NUMBER: US 09/270,957

NUMBER OF SEQ ID NOS: 112

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 8

LENGTH: 602

TYPE: PRT

ORGANISM: Bacillus sp.

US-10-364-649-8

Query Match 31.0%; Score 929; DB 14; Length 602;
 Best Local Similarity 35.6%; Pred. No. 1.5e-68;
 Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy 1 MVRPQRNKKRIFILLGVNMLEV-----TSKDRPIAVGSWNE--QYODLC 44
 Db 1 MLYPINTETRGVDFDLNGVMNFKLDYKGLKEKWESEKLTDTISMAVPSYNDIGVTEIR 60
 Qy 45 YEEGPFYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKYGENHIEYLPEFVDYTGK 104
 Db 61 NHIGVWYEREFYPAVLKQRIYLRFGSATHKALIVYNGELVVEHKGFLPEAEIINS 120
 Qy 105 VKSGENELRVVVENLKVGFSPSKVPDSGTHVGFSGS-----FPANEDFFPYG 154
 Db 121 LRDGMNRTVAVDNL-----DSTLPVGLYSEHREGLGKVIIRNKNPNDFFNYA 170
 Qy 155 GIIRPVLIETDHAIRLDIWDTSSEPEKKLGKVKKIVSEEAAGQEMTIKLG--EE 212
 Db 171 GLHRPVKIYTPPTVEEDISVTDENG---TGTVTYVDFQ----GKAETVKSVDDE 223

Qy 213 KKIIRTSNRFVEGEFLEENAFMSLEDPIYPLKVELEKDEYTLDI-----GIRTI5MDEK 267
 Db 224 GKVVASTEGISGNVEIPNVILMEPLNTLYXQIKVELVNDGLTIDVYEEBPGVRYEVNG 283
 Qy 268 RLYANGKPVFLKPGFKHEEPVLGGQTFYPLMKIDFNLLKWINANSFRSHYPSSEMD 327
 Db 284 KFLINNKPFYFGKGHEDTPINGRGFNEASVMDFNILKMGANSFRTHAYPSSELMR 343
 Qy 328 LADRLGIIVDEAPHVGTIRHYN-----PETOKIA-----EDNIRMIDR 368
 Db 344 LADREGLVVIDETPAVG---HLNFMATTGLGEGSERVSTWEXIKTFEHHQDVLRELVS 400
 Qy 369 HKHPSVIMSVANPEPSNHPDAEGFPKALYETANEMD--RTRPVVMVMDADERTRDV 427
 Db 401 DKHPSVIMSVANPEPSNHPDAEGFPKALYETANEMD--RTRPVVMVMDADERTRDV 458
 Qy 428 ALKYPDIVCNRYGYWYIYQRIEGLQALEKDIIEIYARH---KPIVTEFGADATA 483
 Db 459 VAEIIVIALNRNGWYFDGDLGAAKVHLR---QEFHANNKRCPCPKIMITEYADTVA 515
 Qy 484 GIHYDPQMFSEBYQALVEKTRIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGV 543
 Db 516 GFHDIDPVMTTEBYQVEYQANHVPDEFENFVGEQAMNPAFATSQGVWRVQGNKGV 575
 Qy 544 TRDRPKLVAAHVRRLMSEV 563
 Db 576 TRDRPKLVAAHVRERWTNI 595

RESULT 7
 US-10-364-649-15
 ; Sequence 15, Application US/10364649
 ; Publication No. US2003022921A1

GENERAL INFORMATION:

APPLICANT: Richard A. Jefferson and Jorge E. Mayer

TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE

FILE REFERENCE: 190106.405C1

CURRENT APPLICATION NUMBER: US/10/364,649

PRIOR FILING DATE: 2003-02-12

PRIOR APPLICATION NUMBER: 10/364,649

PRIOR FILING DATE: 2003-02-12

PRIOR APPLICATION NUMBER: US 09/270,957

NUMBER OF SEQ ID NOS: 112

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 15

LENGTH: 602

TYPE: PRT

ORGANISM: Bacillus sp.

US-10-364-649-15

Query Match 31.0%; Score 929; DB 14; Length 602;
 Best Local Similarity 35.6%; Pred. No. 1.5e-68;
 Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy 1 MVRPQRNKKRIFILLGVNMLEV-----TSKDRPIAVGSWNE--QYODLC 44
 Db 1 MLYPINTETRGVDFDLNGVMNFKLDYKGLKEKWESEKLTDTISMAVPSYNDIGVTEIR 60
 Qy 45 YEEGPFYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKYGENHIEYLPEFVDYTGK 104
 Db 61 NHIGVWYEREFYPAVLKQRIYLRFGSATHKALIVYNGELVVEHKGFLPEAEIINS 120
 Qy 105 VKSGENELRVVVENLKVGFSPSKVPDSGTHVGFSGS-----FPANEDFFPYG 154
 Db 121 LRDGMNRTVAVDNL-----DSTLPVGLYSEHREGLGKVIIRNKNPNDFFNYA 170
 Qy 155 GIIRPVLIETDHAIRLDIWDTSSEPEKKLGKVKKIVSEEAAGQEMTIKLG--EE 212
 Db 171 GLHRPVKIYTPPTVEEDISVTDENG---TGTVTYVDFQ----GKAETVKSVDDE 223

QY 213 KKIITSNRFVGEFLENNARFWSLEDPYLPKXVELEKDEYTLDI-----GIRTI5WDEK 267
DB 224 GKVAVSTGSLGNVEIPNVILMBEPLNTLYQIKVELVNDGLTIDYEEBFGVRIYEVNDG 283
QY 268 RLYLNGKRVFLKFGKHEEPVLGOGTFYPLMIKDPNLKWJNANSFRTSHYPYSEEMWD 327
DB 284 KFLINNKFPYFGKFGHEDTPIINGRGFNENASVMDFNILKMGANSFRTAHYPYSEELMR 343
QY 328 LADRLGILVIDEAPVHGITRHYN-----PEQKIA-----EDNIRRMIDR 368
DB 344 LADRLGILVIDETPAVG---HLNFMATTGLGEGSERVSTWERTFEHQVLELVS 400
QY 369 HKNHPSVIMWSVYANPESENHPDAEGFKALYETANEMD-RTRPVYVMSMDAPDERTDY 427
DB 401 DKNHPSVIMWSVYANPESENHPDAEGFKALYETANEMD-RTRPVYVMSMDAPDERTDY 458
QY 428 ALKYPDIVCNRYRYMYIYQGRIEGLQALEKDI BELYARH---KPIFVTEFGADAL 483
DB 459 VAEILDVIALNRKYNMYPDGGDLBAKVHLR---QEFHAMNKRCKGKPMITEYGADIVA 515
QY 484 GIHYDPQMFSEBEXQALVEKTRILLKKDYIIGTHVAFADPKTPQNVRRPILNHKGVF 543
DB 516 GFHDIDPVMFTEBEXQALVEKTRILLKKDYIIGTHVAFADPKTPQNVRRPILNHKGVF 575
QY 544 TRDRPKLVAHYLRRLMSEV 563
DB 576 TRDRPKLVAHYLRRLMSEV 595

RESULT 8
US-10-120-145-4

Sequence 4, Application US/10120145
Publication No. US20030157684A1
GENERAL INFORMATION:
APPLICANT: Jefferson, Richard A.
APPLICANT: Kiliian, Andrzej
APPLICANT: Keese, Paul Konrad
TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
FILE REFERENCE: 190106.405
CURRENT APPLICATION NUMBER: US/10/120.145
PRIOR FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 618
TYPE: PRT
ORGANISM: Bacillus sp.
US-10-120-145-4

Query Match 31.0%; Score 929; DB 14; Length 618;
Best Local Similarity 35.6%; Pred. No. 1.6e-68;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
QY 1 MVRPQRNKRIFILLNGVWNLV-----TSKDRPIAVGSWNE--QYODLC 44
DB 17 MLYPINTETRGVFDLNGVWNLV-----TSKDRPIAVGSWNE--QYODLC 76
QY 45 YEEBGFYKTFYVYKLSQKHIRLYPAVAVTDCVFLNGEKVGENHIEYLPEFVDTGK 104
DB 77 NHIGVYEREFYVPAVILKQRIYLRFGSATHKALVYVNGELVYEHKGGFLPFEALINNS 136
QY 105 VKSGENELRVVYENRLKVGFPSPKVPDGGTHVGFSGS-----FPANDFPFGGII 154
DB 137 LRDGNRVTVAVNDNL-----DSTLPVGLYSERHBEGLGKVI RKNPNDFPNFYA 186
QY 155 GIIRPVLIETFDHARILIDWDTSESEPEKKLGKVKVIEVSEAVQOEMTIKLG--DEE 212
DB 187 GLHHRPKLYTTPFYVEDISVTDENGP---TGIVTYIVDPQ---GAEIVYKVSVDDE 239

QY 213 KKIITSNRFVGEFLENNARFWSLEDPYLPKXVELEKDEYTLDI-----GIRTI5WDEK 267
DB 240 GKVAVSTGSLGNVEIPNVILMBEPLNTLYQIKVELVNDGLTIDYEEBFGVRIYEVNDG 299
QY 268 RLYLNGKRVFLKFGKHEEPVLGOGTFYPLMIKDPNLKWJNANSFRTSHYPYSEEMWD 327
DB 300 KFLINNKFPYFGKFGHEDTPIINGRGFNENASVMDFNILKMGANSFRTAHYPYSEELMR 359
QY 328 LADRLGILVIDEAPVHGITRHYN-----PEQKIA-----EDNIRRMIDR 368
DB 360 LADRLGILVIDETPAVG---HLNFMATTGLGEGSERVSTWERTFEHQVLELVS 416
QY 369 HKNHPSVIMWSVYANPESENHPDAEGFKALYETANEMD-RTRPVYVMSMDAPDERTDY 427
DB 417 DKNHPSVIMWSVYANPESENHPDAEGFKALYETANEMD-RTRPVYVMSMDAPDERTDY 474
QY 428 ALKYPDIVCNRYRYMYIYQGRIEGLQALEKDI BELYARH---KPIFVTEFGADAL 483
DB 475 VAEILDVIALNRKYNMYPDGGDLBAKVHLR---QEFHAMNKRCKGKPMITEYGADIVA 531
QY 484 GIHYDPQMFSEBEXQALVEKTRILLKKDYIIGTHVAFADPKTPQNVRRPILNHKGVF 543
DB 532 GFHDIDPVMFTEBEXQALVEKTRILLKKDYIIGTHVAFADPKTPQNVRRPILNHKGVF 591
QY 544 TRDRPKLVAHYLRRLMSEV 563
DB 592 TRDRPKLVAHYLRRLMSEV 611

RESULT 9
US-10-120-145-8

Sequence 8, Application US/10120145
Publication No. US20030157684A1
GENERAL INFORMATION:
APPLICANT: Jefferson, Richard A.
APPLICANT: Kiliian, Andrzej
APPLICANT: Keese, Paul Konrad
TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
FILE REFERENCE: 190106.405
CURRENT APPLICATION NUMBER: US/10/120.145
PRIOR FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 607
TYPE: PRT
ORGANISM: Bacillus sp.
US-10-120-145-8

Query Match 30.8%; Score 925; DB 14; Length 607;
Best Local Similarity 35.7%; Pred. No. 3.3e-68;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;
QY 4 PQNRKKRIFILLNGVWNLV-----TSKDRPIAVGSWNE--QYODLCYEE 47
DB 9 PINTETRGVFDLNGVWNLV-----TSKDRPIAVGSWNE--QYODLCYEE 68
QY 48 GPFTYKTFYVYKLSQKHIRLYPAVAVTDCVFLNGEKVGENHIEYLPEFVDTGKYS 107
DB 69 GYVYEREFYVPAVILKQRIYLRFGSATHKALVYVNGELVYEHKGGFLPFEALINNSIARD 128
QY 108 GENELRVVYENRLKVGFPSPKVPDGGTHVGFSGS-----FPANDFPFGGII 157
DB 129 GMRNVTVAVNDNL-----DSTLPVGLYSERHBEGLGKVI RKNPNDFPNFYA 178
QY 158 RPYVLIETFDHARILIDWDTSESEPEKKLGKVKVIEVSEAVQOEMTIKLG--DEEKKI 215

```

Db 179 RPVKIYTPPTFYVEDISVTDFNCP---TGTVTYTVDFQ-----GKAEIVKSVVDEBGKV 231
QY 216 RTSNRFESEFILLNARFWSLEDDPYLPYLKVELEKDEYTLDI-----GIRTIISWDEKRLY 270
Db 232 VASTEGLSGIVNEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEBPGRVATVEVNDGKFL 291
QY 271 LNKGPVFLKGFGEHFEFPVLGQGTFFPLMKDPLNLKWINANSFRTSHYSESEWDLAD 330
Db 292 INNKPFYFGFGHEDTLPINGRGNFNASVMDPLNLKIGANSFRTHAHYSESEWDLAD 351
QY 331 RLGIIVIDEAPHVIGITRYHYN-----DETOKIA-----EDNIRMDIRHKN 371
Db 352 REGVLVIDETPAVGV---HLNFMAITGLGSGSERVSTWEEKIRTEBHODVLRBELVSDKN 408
QY 372 HPSVYMSVANPESSNHPDAEGFKALYETANEMD--RTRPVWVMSMDADDERTRDVALK 430
Db 409 HPSVWMSIANEATEEGAYEYEFKPLVELTELDPQKRPVTVLFVMAIPE--TDKVAE 466
QY 431 YPDIVCNRYGYWYIYQRIEELQALEKDIIEELVARRH---KPIFVTEFGADALAGIH 486
Db 467 LIDVIALNRYNGWYFPGDGLDEAKVHLR---QEFHANNKRCPOGKPIITTEYGADTVAGFH 523
QY 487 YDPPQMFSEBYQALVEKTRILLLKKDYIIGTHVMAFADFKTPQNVRRPILNKHGVFTRD 546
Db 524 DIDPVMTFEEYQVEYVQANHVVFEDEFENFVGEQAMNFADFATISQGVWRVQGNKKGVFTRD 583
QY 547 RQPKLVAVHLRLMSEV 563
Db 584 RKPKLAHVFRERWTNI 600

```

RESULT 10
US-10-364-649-18

```

; Sequence 18, Application US/10364649
; Publication No. US2003022921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-18

```

Query Match 30.8%; Score 925; DB 14; Length 607;
Best Local Similarity 35.7%; Pred. No. 3.3e-68;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

```

QY 4 PQRNKRPFILLNGVNWLEV-----TSKDRPIAVGSMNE--QYQDLCYEE 47
Db 9 PINTETRGVFDLNGVWNFKLDYKGLEBKWYESKLTDTISMAVSSYNDIGVTKEINHI 68
QY 48 GPFYKTFYVPKLSQKHRLYFAAVNTDCEVFLNKGKGENHIEVLPEVVDVTKVKS 107
Db 69 GYVWYEREFVPAVLKQRIYLRFGSATHKAIYVNGELVHEHKGFLPFEAEIINSLRD 128
QY 108 GENELRVVENRLKVGSPSKVPSGTHVTFPGS-----PPANPDFPYGGII 157
Db 129 GNRKVTVAVDNL-----DSTLPVGLSERHEBGLGKVIIRKNKPFDFNFVAGLH 178
QY 158 RPYLIEFTDARILDIWDTSESEPEKKGKVKVKEVSEAVGQEMTILG--EEBKTI 215
Db 179 RPVKIYTPPTFYVEDISVTDFNCP---TGTVTYTVDFQ-----GKAEIVKSVVDEBGKV 231

```

```

QY 216 RTSNRFESEFILLNARFWSLEDDPYLPYLKVELEKDEYTLDI-----GIRTIISWDEKRLY 270
Db 232 VASTEGLSGIVNEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEBPGRVATVEVNDGKFL 291
QY 271 LNKGPVFLKGFGEHFEFPVLGQGTFFPLMKDPLNLKWINANSFRTSHYSESEWDLAD 330
Db 292 INNKPFYFGFGHEDTLPINGRGNFNASVMDPLNLKIGANSFRTHAHYSESEWDLAD 351
QY 331 RLGIIVIDEAPHVIGITRYHYN-----DETOKIA-----EDNIRMDIRHKN 371
Db 352 REGVLVIDETPAVGV---HLNFMAITGLGSGSERVSTWEEKIRTEBHODVLRBELVSDKN 408
QY 372 HPSVYMSVANPESSNHPDAEGFKALYETANEMD--RTRPVWVMSMDADDERTRDVALK 430
Db 409 HPSVWMSIANEATEEGAYEYEFKPLVELTELDPQKRPVTVLFVMAIPE--TDKVAE 466
QY 431 YPDIVCNRYGYWYIYQRIEELQALEKDIIEELVARRH---KPIFVTEFGADALAGIH 486
Db 467 LIDVIALNRYNGWYFPGDGLDEAKVHLR---QEFHANNKRCPOGKPIITTEYGADTVAGFH 523
QY 487 YDPPQMFSEBYQALVEKTRILLLKKDYIIGTHVMAFADFKTPQNVRRPILNKHGVFTRD 546
Db 524 DIDPVMTFEEYQVEYVQANHVVFEDEFENFVGEQAMNFADFATISQGVWRVQGNKKGVFTRD 583
QY 547 RQPKLVAVHLRLMSEV 563
Db 584 RKPKLAHVFRERWTNI 600

```

RESULT 11
US-10-364-649-28

```

; Sequence 28, Application US/10364649
; Publication No. US2003022921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-28

```

Query Match 30.8%; Score 925; DB 14; Length 615;
Best Local Similarity 35.7%; Pred. No. 3.4e-68;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

```

QY 4 PQRNKRPFILLNGVNWLEV-----TSKDRPIAVGSMNE--QYQDLCYEE 47
Db 9 PINTETRGVFDLNGVWNFKLDYKGLEBKWYESKLTDTISMAVSSYNDIGVTKEINHI 68
QY 48 GPFYKTFYVPKLSQKHRLYFAAVNTDCEVFLNKGKGENHIEVLPEVVDVTKVKS 107
Db 69 GYVWYEREFVPAVLKQRIYLRFGSATHKAIYVNGELVHEHKGFLPFEAEIINSLRD 128
QY 108 GENELRVVENRLKVGSPSKVPSGTHVTFPGS-----PPANPDFPYGGII 157
Db 129 GNRKVTVAVDNL-----DSTLPVGLSERHEBGLGKVIIRKNKPFDFNFVAGLH 178
QY 158 RPYLIEFTDARILDIWDTSESEPEKKGKVKVKEVSEAVGQEMTILG--EEBKTI 215
Db 179 RPVKIYTPPTFYVEDISVTDFNCP---TGTVTYTVDFQ-----GKAEIVKSVVDEBGKV 231

```

```

RESULT 12
US-09-862-660-2
; Sequence 2, Application US/09862660
; Publication No. US20030003562A1
; GENERAL INFORMATION:
; APPLICANT: Ruseell, William
; TITLE OF INVENTION: LACTOBACILLUS BETA-GLUCURONIDASE AND DNA ENCODING THE SAME
; FILE REFERENCE: 5051.514
; CURRENT APPLICATION NUMBER: US/09/862,660
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/206,372
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patencin version 3.0
; SEQ ID NO 2
; LENGTH: 598
; TYPE: PRN
; ORGANISM: Lactobacillus gasseri
US-09-862-660-2

```

```

QY      278  LKGECKHEEPPVLGGCTGYPLMIKQENLKLKWTNANSPRSHPYSEEWLIDLDRGILVI 337
Db      233  LKGFPEKHEDFVNLGAUVNBSTIKRDIKDYECQKWIAGANCFRSHRYPAEWYQVADKTGFLII 352
QY      338  DEAPVIGITRYIYN-----PETOKIAEDIRIMIDRKHHPSVIAMS 379
Db      353  DEVPAGVGNRSTINTNLANTNSNQSHFPASKTYVPELKYKHGQIKEMIDBDQGNHPSVIAMS 412
QY      360  VANEESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDADESTRDVALKYFDIVCN 438
Db      413  LFNEPESTTOESYDYFKDIFAFARKLBDQGNREPYTGLWVGSQPK--VDKLHPLCDFVCIN 470
QY      439  RYGVGYIYQG-RIEBGLQALEKDIEELY-ARRKRKILPTWEPADNLAGIHNDPQMFSE 496
Db      471  RYGVYVAGGPEITVNAKKMLEDLGMQNLKINKPFVFEFPADTLSSHRLPDEKMSOE 530
QY      497  YQALVEKTIIRLLKKDYIIGTHVAFADFKTPQVWRPRLNHHGVETRDROPKLVANHL 556
Db      531  YQNEVGYWYFDIPIKKYPIFCGLVWNPADFKTISEGIMRVGNDKGIFTDRREPXDIAFTL 590
QY      557  RRLMSEV 563
Db      591  KKRWQOL 597

RESULT 13
US-10-673-935-2
; Sequence 2, Application US/10673935
; Publication No. US20040091922A1
; GENERAL INFORMATION:
; APPLICANT: Ruseell, William
; APPLICANT: Klashammer, Todd
; TITLE OF INVENTION: LACTOBACILLUS BETA-GLUCURONIDASE AND DNA ENCODING THE SAME
; FILE REFERENCE: 5051.514DV
; CURRENT APPLICATION NUMBER: US/10/673,935
; PRIOR FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US 09/862,660
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/206372
; PRIOR FILING DATE: 2000-05-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Lactobacillus gaseeri
US-10-673-935-2

Query Match      30.6%; Score 919.5; DB 15; Length 598;
Best Local Similarity 34.1%; Pred. No. 9,4e-68;
Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;

QY      4  PQRNKKRPIILINGVWNLVET-----SKDRP-----IAPGSWNE--QYODLCYEE 47
Db      7  PIQNKYRNTLNGWQGETDPSNSVGLDEGNMKELPDDEBMPVPTGPAELTTKRDKRYTT 66
QY      48  GPFTYKTTFFYPKXLSQGHIRLPAVAUNTCEVFLNGEKVGNHIEYLPEEVDVTGKVS 107
Db      67  GDFWYQXKDFPIPSFLKKELEYIRFGSVYTHRAKFVINGHAEVQGHGEGFLPFGVYKINYYI 126
QY      108  GE-NELRVYVENRLKVGFP---SKVPSGHTVGFPGSPFRPANDDFPYGGIIRPVULE 163
Db      127  DQNRKVTYLVNNEISEKAIPOGTETILDNQ-----KLAQPYDFPFYSGIMRWVILL 179
QY      164  FTDHARILDIWDTSESEPEKKLGKVKYKIEVSEAVQC-EMTKIKGSEBKKIRTSNRFV 222
Db      180  ALPQSQINPFLN-----YQLANKKATITYYNIEANNAAEKXVTLFPNQKXEVACATSKN 232
QY      223  EGEFLLENARWLSLEDPIILYPLUKVLEK----DEYTLDIGIRTSMBEKRLYLNGKPYF 277
Db      233  TTSLLTIKNPHLMSPDNDPSYKIKIEMEDDKGTVDYEYTKIGIRTVKIVVNDKILLNNHPYI 292

```

QY 278 LKFGKHEEPVUGGTFPLMKDNLKWINANSRTRSHYSEBMDLADRLGLV 317
 DB 293 LKFGKHEEDNNVKGAVNESIIRKDYECMKIGANCRRSSHYAEVQYAKYGEPLI 352
 QY 338 DEAPHVITRYHN-----DETOKIADNIRRMIDRKHNSPYLWMS 379
 DB 353 DEAPANGKRSITNPLANTWMSNOSHFAKSTVEBELKKEHOLEKEMIDRQRHPSYIAMS 412
 QY 380 VANEPESNHDAGEFKALYETANEMD-RTRPVVWMSMDAPERTDVALKYFDIYCVN 438
 DB 413 LFNEBESTQESYDYFDKIDPAFARKLDPQRNRYTGLTWMSGPK--VDKLPDCDFVCLN 470
 QY 439 RYXGWYIYOG-RIEBGLQALEKDIIEEL-ARHKRPIFTEFGADATAGIHYPDPMFSEE 496
 DB 471 RYXGWYVAGGPELVNNAKMLEDELQWQNTKLNKPEVTEFGADTSSSHRLPDEMMSQS 530
 QY 497 YQAEIVKTRRLKLLKDYIIGTHWAFADKTPQNVRRPILNHKGFTRDQKPLVAHV 556
 DB 531 YQAEVYQMYFDIPKYPFICGELVWNPADKRTSGIRVAGNDKGIETRDREPDIATFL 590
 QY 557 RRLMSEV 563
 DB 591 KKGWQL 597

RESULT 14
 US-10-421-175-4
 ; Sequence 4, Application US/10421175
 ; Publication No. US2003021941A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Podaskoff, Gregory
 ; APPLICANT: Watson, Gordon
 ; APPLICANT: Couto, Linda B.
 ; APPLICANT: Yang, Bin
 ; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
 ; FILE REFERENCE: 0800-0021
 ; CURRENT APPLICATION NUMBER: US/10/421,175
 ; CURRENT FILING DATE: 2003-04-22
 ; PRIOR APPLICATION NUMBER: US/09/715,858
 ; PRIOR FILING DATE: 2000-11-14
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 648
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-421-175-4

Query Match 30.0%; Score 899.5; DB 14; Length 648;
 Best Local Similarity 35.9%; Pred. No. 5e-66;
 Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;
 QY 1 MVRPQRKKRFFILINGVNLV-TSKDR-----PIAVGSMNEQ 39
 DB 27 MLFPKESRSRLKXLDGDMHFRALDSNNRLOGFQGWYQRLRSRGPVLDMPVPSFNDI 86
 QY 40 YQDLCTYEE--GPFYKTTFFYPKLSQ--KHILYPAVNTDEVFLNGKVGSENHLEY 94
 DB 87 TQEAALDFTIGWVYERAILPRRWTDPTDMRVVLRINSAYIVAVVWNGIHVHEHGGH 146
 QY 95 LPEFVDVYKYGKSG---ENELRVVVENLTKVGFPSKVP-----DSGHTVGFGSGF 143
 DB 147 LPEFADISKLVSQGPLTTCRITITAINNTLT---PHILPQGTIYKTDTSMPKGYF--V 200
 QY 144 PPAFDFPFGGIRPVLIEFTDARLIDIWDTSESEPEKGLGVK--VKIEVSEAVG 201
 DB 201 QDTSPDFNFVAGHRSVLYTPTTYIDITVIN---VEQDIGVYTWISVQSGSHF-- 255
 QY 202 QEMTKIGEEBKJRTSNRFVGEFIELENARFV---SLEDP-YLYPLKXV-----ELEKD 251
 DB 256 -QLREVQLDDEGGKVAVHGTGNGOQLQVPSANIMWPMYLMHHRPAWMSLEVAVTTTESVTD 314

QY 252 EYTLDIGRTISMDKRLYLNGKFPVYLGKFGHEEPVUGGTFPLMKDNLKWINA 311
 DB 315 YTLPLIGRTVATVTSKFLINCKPFPYFQVGNHGEDSDIRGKGFDPMLVKDNLKWLGA 374
 QY 312 NSFRTRSHYSEBMDLADRLGLVDEAPHVIT-----RYHNPETOKIAEDN 361
 DB 375 NSFRTRSHYSEBMDLADRLGLVDEAPHVIT-----RYHNPETOKIAEDN 429
 QY 362 IRMRIDRKHNSPYLWMSVANEPESNHDAGEFKALYETANEMDRTRVVWMS--MMDA 419
 DB 430 VR----RDKNHVAVVWMSVANEPSSALKPAAVYFKTLITHTYALDITREVTFVSNAKYDA 485
 QY 420 PERTEDVALKFDIYCVNRYXGWYIYQGRIEBGLQALEKDIIEELYARHKRPIFTEFGA 479
 DB 486 -----DLGAPYDVYCVNYSFSWHDYGHLEIVQPLNSQENMWKTHQKPILOSEYGA 539
 QY 480 DALAGIHYPDPMFSEBYQAEIVKTRRL--LKDYIIGTHWAFADKTPQNVRRPIL 537
 DB 540 DALPGIHEDPPRMFSEBYQAEIVENYHSLDQKRKRYVGEILWNPADPMWQSPURVIG 599
 QY 538 NHKGFTRDQKPLVAHV-RLM 560
 DB 600 NKGIFTRQRPKTSFILLERYW 623

RESULT 15
 US-09-118-276-12
 ; Sequence 12, Application US/09118276
 ; Patent No. US20010011381A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BABYCHUK, ELENA;
 ; APPLICANT: KUSHNIR, SERGEI;
 ; APPLICANT: DE BLOK, MARC;
 ; APPLICANT: INZE, DIRK
 ; TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
 ; TITLE OF INVENTION: CELL DEATH IN EUKARYOTIC CELLS
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SIXBEY, FRIEDMAN, LEBDOM, & FERGUSON
 ; STREET: 8180 GREENBORO DRIVE, SUITE 800
 ; CITY: MCLEAN,
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22102
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3-1/2" DISKETTE
 ; COMPUTER: IBM-COMPATIBLE
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/118,276
 ; FILING DATE: 17-JUL-1998
 ; PRIOR APPLICATION DATA: NONE
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
 ; REGISTRATION NUMBER: 31,196; 43,077
 ; REFERENCE/DOCKET NUMBER: 6201-0003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 790-9110
 ; TELEFAX: (703) 883-0370
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1010 RESIDUES
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: SINGLE
 ; TOPOLOGY: LINEAR
 US-09-118-276-12

Query Match 29.4%; Score 882; DB 9; Length 1010;
 Best Local Similarity 34.0%; Pred. No. 2.7e-64;
 Matches 206; Conservative 109; Mismatches 229; Indels 62; Gaps 11;
 QY 1 MVRPQRKKRFFILINGVNLV-----TSKDRPIAVGSMNEQYQDLCY 45


```
Db 408 MVRPVEPTREIKLKDGLMAFSLDRENCIDIDRMWESALQESRAIAVPGSFNDQFADADI 467
Qy 46 EE--GPFTYKKTFFVYPKXLSQKHIRLYPAVNTDCEVFLNGEKVGENHIEYLPEEVDYTG 103
Db 468 RNYAGNVMYQREVEFIPKGMAGQRIVLRFDAVTHYGVKVMWNNQEVMEHOGGYTPFEADYTP 527
Qy 104 KYKSGEN-ELRVVVENRLKVGFPFSKV---PDSGTHYGFPGSFPPANPDPFPGYIIRP 159
Db 528 YVIAKGSVRITVCVNNELMQITPPGMVITTDENGKKQSYF-----HDFPNYAGIHS 580
Qy 160 VLIBFTDARILDIWDTSSEPEKKLGKVKKIEVSEAVQEWITKLGEEKKIRTSN 219
Db 581 VMLYTPNTWDDITVTVTHVAQ---DCNHASVDMQV---VANGDYSVELRDAADQOVATG 634
Qy 220 RPEGEFLENNARFWSLEDPIYPL---KVELEKDEYTLDIGRTISMDEKRLYLNKP 275
Db 635 QGTSGLQVNPMLMQPEGYLCELCTAKSQTECDIYPLRVGIRSVAVKGEOPLINHKP 694
Qy 276 VFLNGFGHSEEPFVVGOGTFYPLMIKDNLKMINANSFRTSHYSEEWDLADRLGITL 335
Db 695 FYTFGFGHEDADLKGKGFNDVLMVHDHALMDWIGANSYRTSHYPAEEMLDWADDEHGIV 754
Qy 336 VIDEAPHYGI-----TRYHNPETOIAEDNIRRMIDRHKNHPSYIM 377
Db 755 VIDETAAVGFNLSLGIFPEAGNKPXELYSEEAVENTOQAHLOALKEILARDKNHPSYIM 814
Qy 378 WSVANEPESNHPDAEGFPKALYETANEMDRTRPVVMVSM--DAPDERTDVALKYFDIV 435
Db 815 WSIANEPTDRPGAREYFAPLAELATRKLDPTFRPITCVNMVFCDAHTDTSIDL---FDVL 870
Qy 436 CVNRYGMYIYQGRIEBGLQALEKIDIEELYARHKKPIFVTERGADALAGIHYDPQMFSE 495
Db 871 CLNRYGMYVOSGDLETAEKYLEKELAMQEKHQPIITTEYGVDTLAGLSMYTDMWSE 930
Qy 496 EYQAEIVEXTIRLLLLKDDYIIGTHVMAFADFPTQNVRRPILNHKGVFTDRDQPKLVAVH 555
Db 931 EYQCAMLDMYHRVPRVSAVVGEOVMNPAFATSQILRVGKNKGIFTRDRKPKSAAPL 990
Qy 556 LRRLMS 561
Db 991 LQKRW 996
```

Search completed: January 24, 2005, 07:55:31
Job time : 85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2005, 22:33:20 ; Search time 854 Seconds

(without alignments)
10382.052 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689
Sequence: 1 atgtctaagaccgcgaacgaa.....gaagactctgagcgaagctt 1689

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003ds:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	3	AAA07937
2	174	10.3	2128	4	AB104147
3	174	10.3	2169	8	AAD50922
4	174	10.3	2169	12	ADJ47502
5	174	10.3	2169	12	ADJ56608
6	172.4	10.2	1956	4	AAD06386
7	172.4	10.2	2191	6	ABK83932
8	172.4	10.2	2191	10	ADD29602
9	172.4	10.2	2191	10	ADG89395
10	172.4	10.2	2191	12	ACF57506
11	172.4	10.2	2191	12	ADJ62864
12	172.4	10.2	2191	12	ADP10326
13	172.4	10.2	3525	4	AAE57574
14	172.4	10.2	7328	4	AAE57573
15	172.8	10.1	3314	2	AA058896
16	166.8	9.9	2472	10	ADB58226
17	166.8	9.9	2472	10	ADB58226
18	159.6	9.4	1947	4	AAD06387
19	145.8	8.6	2109	10	ADG76418
20	142.2	8.4	23078	4	AAE59508
21	142.2	8.4	23078	8	ACF64437

22	141.4	8.4	1887	2	AAZ23826	Aax23826 Bacillus
23	133	7.9	1820	3	AAA07958	Aaa07958 Staphyloc
24	132.6	7.9	1812	3	AAA07939	Aaa07939 Salmonell
25	130.4	7.7	1888	3	AAA07938	Aaa07938 Codon-opt
26	114	6.7	1814	10	ADCC01686	Adc01686 Enterobac
27	114	6.7	3451	8	ABT16609	Abt16609 Artificia
28	114	6.7	3451	10	ACC44711	Acc44711 Plasmid p
29	114	6.7	14627	8	ABT16610	Abt16610 Artificia
30	114	6.7	14627	10	ACC44712	Acc44712 Plasmid p
31	110.8	6.6	1809	6	ABK87080	Abk87080 uidA DNA
32	110.8	6.6	1812	4	AAH22503	Aah22503 PCR gener
33	110.8	6.6	1812	6	AAD29543	Aad29543 uidA DNA
34	110.8	6.6	1812	6	ABE61371	AbE61371 GUS DNA #
35	110.8	6.6	1812	8	ABT16606	Abt16606 Artificia
36	110.8	6.6	1812	10	ACC44709	Acc44709 E. coli b
37	110.8	6.6	2001	4	AAD19830	Aad19830 Beta-gluc
38	110.8	6.6	2001	6	ABK15667	Abk15667 Glucathi
39	110.8	6.6	2186	3	AAZ38598	Aaz38598 Dysfuncti
40	110.8	6.6	2457	1	AAH81451	Aah81451 Sequence
41	110.8	6.6	2673	1	AAH81449	Aah81449 Sequence
42	110.8	6.6	2725	4	AAD19835	Aad19835 Promoter-
43	110.8	6.6	2730	4	AAD19833	Aad19833 Promoter-
44	110.8	6.6	3035	2	AAV37748	Aav37748 PAT1 gene
45	110.8	6.6	3169	2	AAZ38397	Aaz38397 pTtc 99 p

ALIGNMENTS

RESULT 1
AAA07937
ID AAA07937 standard; DNA, 1689 BP.
XX
AC AAA07937;
XX
DT 26-JAN-2001 (first entry)
XX
DE Thermotoga maritima beta-glucuronidase gene.
XX
KW Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;
KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KW transgenic insect; marker; glucuronide detoxification; de.
XX
OS Thermotoga maritima.
XX
PN WO200055333-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US007107.
XX
PR 17-MAR-1999; 99US-00270957.
XX
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Mayer JE;
XX
DR WPI; 2000-64705/62.
XX
PS P-PSDB; AAB28406.
XX
PT Novel microbial beta-glucuronidase genes and gene products used as
PT reporter/effector molecule, as diagnostic tool, in positive selection, to
PT target molecules to specific cells and to detect and track linked genes.
XX
PS Claim 2; Fig 4; 116pp; English.
XX
CC The present sequence encodes a microbial beta-glucuronidase (GUS)
CC protein. GUS genes were obtained from six different genera:
CC Enterobacter/Salmonella, Pseudomonas, Salmonella, Staphylococcus and
CC Thermotoga. Microbial GUS can be used as a reporter/effector molecule for
CC transgenic constructions and in in vitro diagnostic applications. It may
CC also be used to generate sentinel plants that serve as bioindicators of
CC environmental status. It may be used to generate transgenic insects for

CC tracking insect populations or to facilitate the development of a
 CC bioassay for compounds that affect molecules critical for insect
 CC development (e.g. juvenile hormone). Secreted GUS may also serve as a
 CC marker for beneficial fungi destined for release into the environment. In
 CC animal systems, secreted GUS may be used to achieve extracellular
 CC detoxification of glucuronides (e.g. toxin glucuronide) and to examine
 CC conjugation patterns of glucuronides. Microbial GUS may also be used in
 CC traditional medical diagnostic assays, for drug testing, pharmacokinetic
 CC studies, bioavailability studies, diagnosis of diseases and syndromes,
 CC following progression of disease or its response to therapy. Microbial
 CC GUS has increased thermal stability, high turnover number and enzymatic
 CC activity. It is highly specific for the substrate and water soluble, and
 CC the substrates are stable

XX Sequence 1689 BP; 526 A; 374 C; 430 G; 358 T; 0 U; 1 Other;

Query Match 99.9%; Score 1688; DB 3; Length 1689;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;

Matches 1689; Conservative 0; Indels 0; Gaps 0;

QY 1 ATGTGAAGCCGCAACGAAACAGAGAGATTATCTTATCTTGAATGAGTTGGAT 60
 DB 1 ATGTGAAGCCGCAACGAAACAGAGAGATTATCTTATCTTGAATGAGTTGGAT 60
 QY 61 CTGGAATACCGCAACGAAACAGAGAGATTATCTTATCTTGAATGAGTTGGAT 120
 DB 61 CTGGAATACCGCAACGAAACAGAGAGATTATCTTATCTTGAATGAGTTGGAT 120
 QY 121 CAGGATCTGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 DB 121 CAGGATCTGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 QY 181 NAACCTTACAAAACACATCAGATTACTTCTGCTGCTGAGAGAGAGAGAGAG 240
 DB 181 NAACCTTACAAAACACATCAGATTACTTCTGCTGCTGAGAGAGAGAGAGAG 240
 QY 241 TTCTCTCAACGAG 300
 DB 241 TTCTCTCAACGAG 300
 QY 301 GTGACGGGGAAG 360
 DB 301 GTGACGGGGAAG 360
 QY 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACCGTGG 420
 DB 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACCGTGG 420
 QY 421 GGAAGTTTCCACCTGCAAACTTGAAGTTCCAGAGGTTGAGATCATTAAGGCT 480
 DB 421 GGAAGTTTCCACCTGCAAACTTGAAGTTCCAGAGGTTGAGATCATTAAGGCT 480
 QY 481 CTGATAGAGTTCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 DB 481 CTGATAGAGTTCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 QY 541 GAACCGGAGAGAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 DB 541 GAACCGGAGAGAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 601 GGACGAGAGATGAGATCAAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 DB 601 GGACGAGAGATGAGATCAAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 661 TTGCTGAGAGGAGAGTTGATCTGAGAAACGCAAGTTCTGAGAGCTTGAAGAT 720
 DB 661 TTGCTGAGAGGAGAGTTGATCTGAGAAACGCAAGTTCTGAGAGCTTGAAGAT 720
 QY 721 CTCTTATCTCTCAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 DB 721 CTCTTATCTCTCAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 QY 781 ACGATCAGCTGGAG 840

DB 781 ACGATCAGCTGGAG 840
 QY 841 TTGGAAGACAGAGAGATTTCCGCTTGGGGCAGGGCAGCTTTTATCATTTGATGA 900
 DB 841 TTGGAAGACAGAGAGATTTCCGCTTGGGGCAGGGCAGCTTTTATCATTTGATGA 900
 QY 901 AAAGACTTCAACCTTCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 DB 901 AAAGACTTCAACCTTCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 QY 961 TACAGTGAAGAGTGGCTGATCTTGGCCGACAGACTCGAATCTTGTGTATGAGAGCC 1020
 DB 961 TACAGTGAAGAGTGGCTGATCTTGGCCGACAGACTCGAATCTTGTGTATGAGAGCC 1020
 QY 1021 CCGCAGCTTGTATCAAGAGTACCACTTCAATCCGAGACTTCAGAGATGAGAGAGCC 1080
 DB 1021 CCGCAGCTTGTATCAAGAGTACCACTTCAATCCGAGACTTCAGAGATGAGAGAGCC 1080
 QY 1081 AACATGAAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 DB 1081 AACATGAAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 QY 1141 GCGAAGCAACAGAGTCCAAACATCCAGACGCGAGAGGTTTCTTCAAGCCCTTATGAG 1200
 DB 1141 GCGAAGCAACAGAGTCCAAACATCCAGACGCGAGAGGTTTCTTCAAGCCCTTATGAG 1200
 QY 1201 ACTGCCATGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 DB 1201 ACTGCCATGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1261 GACGAG 1320
 DB 1261 GACGAG 1320
 QY 1321 TACGCTGTGTATCTTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 DB 1321 TACGCTGTGTATCTTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1381 ATGGAAGAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 DB 1381 ATGGAAGAGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1441 GCGATAGCTGAGATCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 DB 1441 GCGATAGCTGAGATCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 QY 1501 CTGCTTGAAGAGAGATCAAGAGCTCTTTTGAAGAGAGAGAGAGAGAGAGAGAG 1560
 DB 1501 CTGCTTGAAGAGAGATCAAGAGCTCTTTTGAAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 1561 TGGGCTTTCAGATTTTAAAGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 DB 1561 TGGGCTTTCAGATTTTAAAGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1621 GGTGTTTTCAG 1680
 DB 1621 GGTGTTTTCAG 1680
 QY 1681 AGTGAAGTT 1689
 DB 1681 AGTGAAGTT 1689

RESULT 2

ABL04147
 ID ABL04147 standard; cDNA; 2128 BP.

XX ABL04147;

AC ABL04147;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 6923.

PS Claim 1, SEQ ID NO 6923; 21pp + Sequence Listing; English.
xx
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB161576-AB130511), expressed DNA
CC sequences (AB101840-AB161175) and the encoded proteins (BAB57737-
CC BAB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WFO at [ftp.wipo.int/pub/published_pot_sequences](http://wipo.int/pub/published_pot_sequences)

Query Match	10.3%	Score 174;	DB 4;	Length 2128;
Best Local Similarity	51.6%	Pred. No. 2.5e-43;		
Matches 492; Conservative	0;	Mismatches 435;	Indels 27;	Gaps 3

Qy	734	AGGTGAACTTTGAAAAAGACGAGTACACTCTGTGACATCGGAATCAGACGATCAGTGGG	793
Db	1104	AGTTGGAAATCTCTGCAGATTAACCTTACCGGCTCCAGGTGGCATACCTTAAGTTGGG	1163
Qy	794	ACGAGAAAGAGGCTCTATCTGAAACGGGAAACCTGCTTTTGTGAAGGGCTTTGGAAAGCAG	853
Db	1164	ACAAAGCAGCGCTGCTGCTGAAATGAAAAACCCCTCTATCTCGGGGSAATTTGACGGCAGC	1222
Qy	854	AGGAATTTCCCGTTCTGGGGCAGGGCACCTTTTATCATTTGATGATPAAAAGACTTCAAC	913
Db	1224	AGGACTCCGATATCCGCGGGAAAGGATGTGATPAACTGCGCTTCTTGCTAAGATTTTAAAC	1282
Qy	914	TTCTGAAGTGATCMAACGCGAATTTCTTTAGGACCTCTCACTATCCTTACAGTGAAGAT	973
Db	1284	TGCTGAAGTGACCTGGAGCCAAATGCATATTCGACCTCTCACTATCCTTATTCGGAAGAT	1344
Qy	974	GGCTGGATCTTGCACCGACAGACTCGGAATCCTTGATGATACGGAAGCCCGCACGTTGGTA	1033
Db	1344	CAATGCAGTTTGCCCAATCAGCATGATGATTAATGATTTTGAACGAAATCCC-----TGCTG	1397
Qy	1034	TCACAAGGTACCACTTACAAATCCCGAGACTCAGAAGATACAGAAAGCAACATPAAAGAA	1093
Db	1398	TCAATATGATATCTTCGAGCGGACAGCTACTGGAGAAATCACATGTCCTCGCTGAGCAAC	1455
Qy	1094	TGATGCACAGACACMAAGAACATCCCAAGTGTATCATATGAGAGTGTGGGAATCGAACAG	1155
Db	1458	TGATTCACCGGACAGAAACCATTCMAAGTGTGTTGCATGTGTGGTGAAGCCAACGAGCCA	1511

RESULT 3
AAD50922
ID AAD50922 standard; DNA; 2169 BP.
XX
XX AAD50922;
AC
XX
XX
DT 02-APR-2003 (first entry)
XX
XX
DE L. mexicana SAp-human beta-GUS-IGF-II chimeric DNA.

KM	lysosome; metabolic disease; lysosomal storage disease; gene therapy;
KM	Gaucher's disease; Pompe's disease; Hurler's syndrome; neuroprotective;
KM	Niemann-Pick's disease; Schindler's disease; mucopolidosi; cystinosis;
KM	Batten's disease; prosopoin; infantile neuronal ceroid lipofuscinosis;
KM	fucosidosis; mannosidosis; antileptemic; insulin-like growth factor 2;
KM	IGF-II; secreted acid phosphatase; SAP; beta-glucuronidase; GUS; human;
KM	chimeric; gene; ds.
XX	
XX	
OS	Homo sapiens.
OS	Leishmania mexicana.
OS	Unidentified.
OS	Chimeric.
XX	
XX	
FH	Key
FT	Location/Qualifiers
FT	1. 2169
FT	/*tag= a
FT	/product= "L. mexicana SAP-human beta-GUS-IGF-II chimeric
FT	protein"
FT	1. .96
FT	/*tag= b
FT	/note= "Leishmania mexicana SAP peptide"
FT	97. .1806
FT	/*tag= c
FT	/product= "Human mature beta-GUS protein"
FT	1807. .1815
FT	/*tag= d
FT	misc_feature

Db	1418	TGGACCCCTCCCGGCTGTGACCTTTGTGAGCACTCTAATCTATGAG-----	1465
Qy	1274	GAGACGTGGCGCTGAAGTACTTGCACATTCGTCTGTGTGAACAGGTACTACGGCTGTGACA	1333
Db	1466	CAGACAAAGGGGGCTCCGTATGTGGATGTGATCTTTTGAACAGCTACTACTTTTGGTATC	1525
Qy	1334	TCCTATCAGGGGAAGATGAGAGAGAACTTCAGCTCTGGAAAAAGACATGAGAGCTCT	1393
Db	1526	ACGACTACGGGCACCTGGAGTTGATTCAGCTGCAGCTGGCCACCAGTTTGAGAACTGTT	1585
Qy	1394	ATGCAAGSCACAGAAAGCCACTTTTGTTCACAGAAATTCGTGGCGAAGCCATGACTGCGCA	1453
Db	1586	ATTAAGATATCAAGAGCCCTTATTTACAGACGGATGTGAGACAGAAACGATTCACAGGGT	1645
Qy	1454	TCCACTACGATCCACCTCCAAATGTTCTCCGAGAGTACCAAGCAGAGCTGTTGAAAG-	1512
Db	1646	TTCAACCAAGATCCACCTCTGTATGTTCACTGAAGAGTACAGAAAGAATCTGTGAGCAGT	1705
Qy	1513	ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGGAACAACGCTGGCCCT	1567
Db	1706	ACCATCTGGGGCTCGGATCCAAAAAGCAGAAATAATGTGGTGGAGACTCATTTGGGATTT	1765
Qy	1568	TTGAGATTTTAAAGTCTCTCAGAAATGTGAGAAACCAATTCCTCAACACAGGGGTGTT	1627
Db	1766	TTTCCGATTTTATATCTATGAAACAGTCACCCGAGAGATGCTGGGGAAATTAATAAGGGGATCT	1825
Qy	1628	TCACAAGAGACAGACAAACCCAAACTCGTGTGCTCATGTACTGAGA 1671	
Db	1826	TCACCTCGGCAAGACAAACCAAAAAGTCAAGGTTTCCTTTGGCA 1869	

RESULT 7
ABK83932 standard; cDNA; 2191 BP.

XX ABE83932;
 AC
 XX
 DT 14-AUG-2002 (first entry)

DE	Human cDNA differentially expressed in granulocytic cells #503.
XX	
XX	Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW	viral infection; parasitic infection; protozoal infection;
KW	fungal infection; sterile inflammatory disease; psoriasis;
KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW	cardiac reperfusion injury; renal reperfusion injury; AIDS;
KW	adult respiratory distress syndrome; inflammatory bowel disease;
KW	Crohn's disease; ulcerative colitis; periodontal disease;
KW	granulocyte activation; chronic inflammation; allergy.
XX	
OS	Homo sapiens.
XX	
PN	WO200228999-A2.
XX	
PD	11-APR-2002.
XX	
PF	03-OCT-2001; 2001WO-US030821.
XX	
PR	03-OCT-2000; 2000US-0237189P.
XX	
PA	(GENE-) GENE LOGIC INC.
XX	
PI	Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;
XX	
DR	WPI; 2002-435328/46.
XX	
PT	Detecting granulocyte activation by detecting differential expression of
PT	genes associated with granulocyte activation, which serves as diagnostic
PT	markers that is useful for monitoring disease states and drug toxicity.
XX	
PS	Claim 1; SEQ ID NO 503; 114pp; English.
XX	
CC	The invention relates to detecting (M1) granulocyte (GC) activation
XX	

CC cancer is indicative of cancer. The polypeptide is selected from any of
CC the polypeptides encoded by the polynucleotides listed in the
CC specification and polypeptides which are at least 70% homologous to the
CC polypeptides. The method of the invention has cytostatic activity, and
CC may have a use in gene therapy. The method is useful in identifying
CC markers specific for one or several types of cancer, depending on the
CC tissue origin, which may be used in numerous diagnostic and prognostic
CC applications as well as cancer type-specific targets for therapeutic
CC intervention. The compounds that modulate the activity of a tumour
CC suppressor gene are useful in the treatment of cancer or as anti-cancer
CC drugs. The present sequence represents a polynucleotide of the invention.
XX
XX Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Qy	1568	TTGCAGATTTTAAACATCTCCAGAAATGTGAGAAACCATTCTCAACCAAGGTTGT	1627
Db	1792	TTGCCGATTTATACCTAGAACAGTCACCGACGAAAGTGTCTGGGAAATAAAAAGGGATCT	1851
Qy	1628	TCACAAGAGACAGACAAACCCAAATCTGTCCTCATGTACTAGAG	1671
Db	1852	TCACCTCGGACAGACAAACCAAAAGTGCAGGCTTCTTTTGCAG	1895

RESULT 9
ADG89395 standard; DNA; 2191 BP.

RESULT 9
ADG89395
ID ADG89395 standard; DNA; 2191 BP.

XX	ADG89395;	
AC		
XX		
DT	11-MAR-2004 (first entry)	
XX		
DE	Cancer detection method related gene #58	

KW ds; cancer; gene expression; estrogen receptor-positive invasive breast cancer.

OS Homo sapiens.

PN WO2003078662-A1.

PD 25-SEP-2003

PF 12-MAR-2003; 2003WO-US007713.

PR 13-MAR-2002; 2002US-0364890P.

XX

PA (GENO-) GENOMIC HEALTH INC.

PI Baker JB, Cronin MT, Kiefer MC, Shak S, Walker MG,

DR WPI; 2003-767536/72.

PT Predicting clinical outcome for a patient diagnosed with cancer comprised
PT determining the expression level of one or more genes, and compared to
PT the amount found in a reference cancer tissue set.

PS Disclosure; SEQ ID NO 343; 198pp; English.

The invention relates to a method of predicting clinical outcome for a patient diagnosed with cancer by determining the expression level of one or more genes, or their expression products, selected from p53BP2, cathepsin B, cathepsin L, Ki67/Mib1, and thymidine kinase in a cancer tissue obtained from the patient, normalized against control gene(s), and compared to the amount found in a reference cancer tissue set. The specification also discloses an array comprising polynucleotides 67, GSTM1, CAG, PR, BRCA3, NME1, SRV, GATA3, TFGC, YB-1, DPYD, GSTM3, RPS6KB1, SFO, CHN1, ID1, BstR1, p27, CCN1, XIAP, CHK2, CICC25B, IGF1R, AKO56599, P13KCA2, TGFBR3, BAG1, CYP3A4, EPCAM, VEGFC, p52, HENT1, WISP1, HNF3A, NFKBIP6, BRCA2, EGFR, TTK, VDR, C12orf1037, PENT1, EPX1, IFI1A, CDH1, H1P1, IGFEP3, CTGB, Her2 and DIABLO, immobilized on a solid surface. The methods are useful for predicting clinical outcome for a patient diagnosed with cancer, classifying cancer, and predicting the likelihood of long-term survival of a breast cancer patient, or a patient diagnosed with invasive breast cancer or with estrogen receptor (ER)-positive invasive breast cancer. This sequence corresponds to a gene sequence whose expression is detected by the method of the invention.

Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match	10.2%	Score 172.4;	DB 10;	Length 2191;
Best Local Similarity	52.1%;	Pred. No. 8e-43;		
Matches 492: Conservative	0;	Mismatches 431;	Indels 21;	Gaps 4

737 TGGACTTGAAAAGAGCGAGTACACTCTGGACATCGGAATCAGACGATCAGCTGGACG 796

```

Db      964 TGGGGCCCTGTCTGACTTCTTACACACTCCTCTGGGGATCCGACTGTGCTGTACCA 1023
Qy      797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAGCAGAG 856
Db      1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTCTCCAGGTGTCAACAAAGCATGAG 1083
Qy      857 AATTCCCGCTTCGGGGCAGGGCACCTTTATCCATTGATGATAAAGACTTCAACCTTC 916
Db      1084 ATGCGGACATCCGAGGGAAGGGCTTCGACTGGCCGCTGCTGTAAGGACTTCAACCTGC 1143
Qy      917 TGAAGTGATCAAGCGCAATCTTTTCAGAGCCTCTCACTATCTTACAGTGAAGTGGC 976
Db      1144 TTGCGTGGCTTGGTGCCAAACCTTTCCGTACACACCACTACCTTATGACAGGAATGA 1203
Qy      977 TGGATCTTGGCCGACAGACTCGAAATCTTGTGATGAGCAAGCCCGCAGCTTGTATCA 1036
Db      1204 TGCAGATGTGTGACCGCTATGGGATGTGTGTCATGATGATGCCGCGTGGCCCTGG 1263
Qy      1037 CAAGGTACCA---CTACATCCGAGACTCGAAGATAGCAAGAGCAACATAGAAGAA 1093
Db      1264 CGTGGCGCACTTCTTCAACAACGTTTCTGTGATCACACATGCAAGGTGATGGAAGAG 1323
Qy      1094 TGATCGACAGACAGAAAGACCATCCAGTGTGATCATGTGGAGTGTGGCGAAGCAAG 1153
Db      1324 TGTGCGGTAGGAGCAAGAACCAACCCCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1383
Qy      1154 AGTCCAACTCCAGACGCGGAGGCTTCTTCAAAAGCCCTTATAGACTGCCAATGAA 1213
Db      1384 CGTCCCACTTGAATCTGTCTGTGCTACTACTTGAAGATGTGTANTGCTCACACCAAACTCT 1443
Qy      1214 TGGATCGAACAGCGCCCGTGTGTGATGAGCATGATGAGACGACCAAGAGAGAAACAA 1273
Db      1444 TGGACCTCTCCCGCTGTGACTTGTGAGCACTCTAACTATGCAG-----1491
Qy      1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTGTGTGAAACAGTACTTACCGCTGTACA 1333
Db      1492 CAGACAAAGGGGGCTCCGATGTGATGTGATGTGATGTGAAACAGTACTTGTGTATC 1551
Qy      1334 TCTATCAGGGAAGATGAGAAAGAGACTTCAAGCTCTGGAAGAAAGACATAGAGACTCT 1393
Db      1552 ACAGCTACGGGCACTGGAGTTGATTCAGCTGACGCTGGCCACCAAGTTGAGAACTGTG 1611
Qy      1394 ATGCAAGGACAGAAAGCCATCTTTGTGCACAGAAATCGGTGGGAGCGCATAGCTGCA 1453
Db      1612 ATAAGAAATATCAAAAGCCATTATTCAGAGCGATATGAGAGCAAAACATTTGCAAGGT 1671
Qy      1454 TCCACTACGATCACTTCAAAATGTTTCCGAAAGATACCAAGACAGCTCGTTGAAAG- 1512
Db      1672 TTCAACGAGATCCACTCTGATGTTCACTGAAAGATACGAAAGAGTCTGCTAGAGAGT 1731
Qy      1513 ACGATCAGGCTCTTTTGAAGAA-----AGACTACATCATCTGGAACAACAGTGGGCT 1567
Db      1732 ACCATCTGGGCTGGATCAAAACGCAAGAAATATGTGTGAGAGCTCAATTTGAAT 1791
Qy      1568 TTGAGATTTTAAAGTCTCTCAGAAATGAGAAAGACCATTTCAACCAAGAGGTGTT 1627
Db      1792 TTGCGCATTTTATATGATGAGAACATGACCGAGAGAGTGTGGGAATTAAGAGGAGATCT 1851
Qy      1628 TCACAAGAGACAGCAACCCAACTGTTGCTCATGTGATGAGA 1671
Db      1852 TCACCTGGGAGAGCAACAAAGAAAGTGCAGGCTTCTTTTGGCA 1895

```

RESULT 10

ACF57306

ACF57506

ACF57506; standard; DNA; 2191 BP.

22-APR-2004 (first entry)

Human beta-glucuronidase DNA SEQ ID NO: 7.

```

XX      KM Protein targeting; angiogenesis inhibitor; fusion gene; cancer; gene; de;
KM      KM antiangiogenic; cyostatic; ophthalmological; antiarthritic;
KM      KM antithrombotic; antiinflammatory; dermatological; immunosuppressive;
XX      KM antiporiatric; antidiabetic; neuroprotective; vasotropic; anorectic.
OS      Homo sapiens.
XX      FN NC0203093303-A1.
XX      PD 13-NOV-2003.
XX      PF 06-MAY-2003; 2003MO-US014243.
XX      PR 06-MAY-2002; 2002US-0380063P.
XX      PA (TEXA ) UNIV TEXAS SYSTEM.
XX      PI Hung M, Lan K, Ou-Yang F, Liu J, Lan K;
XX      DR WPI; 2004-053038/05.
XX      PT New polypeptide comprising an antiangiogenesis polypeptide region linked
XX      PT to a therapeutic protein or a diagnostic protein, useful for diagnosing
XX      PT or treating angiogenesis-dependent diseases e.g. cancer, lupus or
XX      PT restenosis.
XX      PS Disclosure; Page 86; 0pp; English.
XX      CC The present invention relates to a polypeptide comprising an
XX      CC antiangiogenesis polypeptide region linked to a therapeutic protein or
XX      CC polypeptide region, or a diagnostic protein or polypeptide region. The
XX      CC polypeptide comprising an antiangiogenesis polypeptide region or the
XX      CC nucleic acid encoding the polypeptide comprising the antiangiogenesis
XX      CC polypeptide is useful for diagnosing or treating angiogenesis-dependent
XX      CC diseases or cancer. Angiogenesis-dependent diseases include age-related
XX      CC macular degeneration, arteriosclerosis, angiodioma, neovascular
XX      CC glaucoma, arteriovenous malformations, nonunion fractures, arthritis,
XX      CC rheumatoid arthritis, lupus, connective tissue disorders, Osler-Weber
XX      CC syndrome, psoriasis, corneal graft neovascularization, pyogenic
XX      CC granuloma, delayed wound healing, retrolental fibroplasia, diabetic
XX      CC retinopathy, scleroderma, granulata, haemangioma, trachoma,
XX      CC haemophilic joints, vascular adhesions, hypertrophic scars, multiple
XX      CC sclerosis, restenosis or obesity. The present sequence is a coding
XX      CC sequence shown in the exemplification of the invention.
XX      SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;
Qy      Query Match 10.2%; Score 172.4; DB 12; Length 2191;
Qy      Best Local Similarity 52.1%; Pred. No. 8e-43;
Qy      Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;
Qy      737 TGGAACTTGAAGAAAGCGAGTACCTTGGACATCGGAATCAAGACATGACTGGAGC 796
Db      964 TGGGGCCCTGTGTCTGACTTCTTACACACTCCTCTGGGGATCCGACTGTGCTGTACCA 1023
Qy      797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAGCAGAG 856
Db      1024 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTTCACGGTGTCAACAGCATGAG 1083
Qy      857 AATTCCCGCTTCGGGGCAGGGCACCTTTATCCATTGATGATAAAGACTTCAACCTTC 916
Db      1084 ATGCGGACATCCGAGGGAAGGGCTTCGACTGGCCGCTGCTGTAAGGACTTCAACCTGC 1143
Qy      917 TGAAGTGATCAAGCGGAATCTTTTCAAGACCTCTCATATCTTACAGTGAAGTGGC 976
Db      1144 TTGCGTGGCTTGGTGCCAAACCTTTCCGTACACACCACTACCTTATGACAGGAATGA 1203
Qy      977 TGGATCTTGGCCGACAGACTCGAAATCTTGTGATGAGCAAGCCCGCAGCTTGTATCA 1036
Db      1204 TGTGCGGTAGGAGCAAGAACCAACCCCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1263
Qy      1037 CAAGGTACCA---CTACATCCGAGACTCGAAGATAGCAAGAGCAACATAGAAGAA 1093

```

```

Db      1264 CGCTGCCGAGTTCTTCAACAAAGTTCTCTGATACCAACATGACAGTGTGAAGAG 1323
Qy      1094 TGATGACACGACACAAACATCCAGTGTGATCATGTGAGTGTGGCAACGAC 1153
Db      1334 TGGTCGTGAGGACAAAGACACCCCGCGTGTGATGTGTCTGTGGCCAAAGAGCTTG 1383
Qy      1154 AGTCCAACCATCCAGACCGCGAGGGTTTCTTCAAAAGCCCTTATGAGACTGGCAATGAAA 1213
Db      1384 GTTCCCACTAGAACTCTGCTGCTACTTGAAGTGTGATTCGCTACACCAATCTCT 1443
Qy      1214 TGGATCGAACACGCCCCGTTGTTCATGTGAGATGATGAGACCGACACGACGAGAACAA 1273
Db      1444 TGGACCCCTCCGCGCTGTGACCTTTGTGAGCACTTAACATGAG----- 1491
Qy      1274 GAGACGTGGCGCTGAAGTACTTGACATCGTGTGTGAACAGGTACTACGGCTGTACA 1333
Db      1492 CAGACAAAGGGGCTCCGTATGTGTGATGTGATCTGTTGAACAGCTACTCTCTGTGATC 1551
Qy      1334 TCTATCAGGAGAGATAGAAAGAGCTTCAGAGCTTGGAAAAACATAGAAAGCTCT 1393
Db      1552 ACGACTACGGGACCTGAGTGTGATTCAGCTGCAGCTGGCCACCGATTGGAAGTGGT 1611
Qy      1394 ATGCAAGGACAGAAAGCCCATCTTGTTCACAGAAATTCGCTGCGACCGCATAGCTGCA 1453
Db      1612 ATAAAGATGAGAAAGCCATTATTCAGAGGAGATGAGACGAAAGATTCAGAGGT 1671
Qy      1454 TCCACTAGATCCACTCAATGTCTCCGAAGAGTACCAAGACAGCTCGTTGAAAAG- 1512
Db      1672 TTCACGAGATCCACTCTGATGTTCACAGAGATACAGAAAGTGTGATGAGCAGT 1731
Qy      1513 ACGATCAGGCTCTTTTGAAGAA-----AGACTACATATCGGAACACACCGTGGGCT 1567
Db      1732 ACCATCTGGGTGTGATCAAAACCCAGAAATATGTGTGGAAGCTCATTTGGAATT 1791
Qy      1568 TTGCAATTTTAAAGTCTCTCAGAAATGTGAGAAAGCCATCTCAACCAAGAGGTGTT 1627
Db      1792 TTGGCGATTTCATGATCTGAACGATCAGCAGACGATGCTGGGGAATTAAGAGGAGTCT 1851
Qy      1628 TCACAAAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
Db      1852 TCACCTCGGACAGACAAACCAAAAGTGACAGCTTCTTTGCGA 1895

```

RESULT 11
ADJ62864
ID ADJ62864 standard; DNA; 2191 BP.
XX
AC ADJ62864;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human beta-glucuronidase gene related to cancer treatment.
XX
KW beta-catenin; bipartite T-cell factor; Tcf; promoter construct; Lef-1;
KW cytoskeletal; beta-catenin activity inhibitor; gene therapy; colon cancer;
KW metastasis; liver; thymidine kinase; prodrug; chemotherapy;
KW radiation therapy; surgery; human; beta-glucuronidase; gene; ds.
XX
OS Homo sapiens.
XX
PN US200328285-A1.
XX
PD 11-DEC-2003.
XX
PF 05-MAY-2003; 2003US-00429802.
XX
PR 03-MAY-2002; 2002US-0377672P.
XX
PA (HUNG/) HUNG M.
PA (KWON/) KWONG K Y.
PA (ZOU/) ZOU Y.
XX

```

PI Hung M, Kwong KY, Zou Y;
XX
XX WPI; 2004-042209/04.
XX
XX Novel viral vector comprising beta-catenin/bipartite T-cell factor-
PT responsive promoter having first and second promoter region linked to
PT target nucleic acid sequence, useful for treat- ing colon cancer.
XX
XX Disclosure: SEQ ID NO 16; 114bp; English.
XX
XX This invention relates to a novel viral vector comprising a beta-
CC catenin/bipartite T-cell factor (Tcf)-responsive promoter construct which
CC contains a first promoter region having a copy of Tcf/Lef-1 binding site
CC operatively linked to a second promoter region, and a nucleic acid
CC sequence, where the first and second promoter regions are operatively
CC linked to target nucleic acid sequence. The invention may be useful for
CC the development of compounds with a cytostatic activity, through action
CC as beta-catenin activity inhibitors, or for gene therapy. The invention
CC may be useful for treating an individual with colon cancer which is
CC metastasised to the liver. The treatment involves administering the
CC vector of the invention where the nucleic acid sequence encodes a
CC therapeutic polypeptide or thymidine kinase, a prodrug and chemotherapy,
CC radiation, surgery or gene therapy to the individual. The present
CC sequence is that of a therapeutic gene which may be used in the vector of
CC the invention for the treatment of an individual with cancer.
XX
XX Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;
SQ
Query Match 10.2%; Score 172.4; DB 12; Length 2191;
Best Local Similarity 52.1%; Pred. No. 8e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

737 TGGAACTGAAAAAGACGATGACTGTGACATCGAATGAAAGATGACGTGAGCG 796
Db TGGGGCTGTGTCTGACTTACACACCTCGTGGGAGCTCGACTGTGCTGACCA 1023
Qy 797 AGAAGAGCTCTATCTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAAGCAGAG 856
Db AGAGCAATTCCTCATCATGAGGAAACCTTTTATTTCCAGGCTGTCAAACAGCATGAGG 1083
Qy 857 AATTCCCGCTTGGGGGACGACCTTTTATTCATGATGATTAAGACTTCAACTTC 916
Db ATCGGATATCCGAGGAGAGGGCTTCGACTGGCCGCTCTGTGAAGACTTCAACCTGC 1143
Qy 917 TGAAGTGATACGCGAATTTCTTCAGACCTCTCACTATCTTTACAGTGAAGTGGC 976
Db TTGCTGTGCTTGTGTCACCAACCTTTCCGTACAGCACTTACCCGTATGAGGAATGGA 1203
Qy 977 TGGATCTGGCCGACAGACTCGGAATCTTGTGATGAGAGAGCCCGACCTGGTATCA 1036
Db TGCAGATGTGACCGCTATGAGGATGTGTGATCATGAGATGTGCCGCGTGGGCTTGG 1263
Qy 1204 TGCAGATGTGACCGCTATGAGGATGTGTGATCATGAGATGTGCCGCGTGGGCTTGG 1263
Qy 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATGACAGAAACAACTTAAGAGAA 1093
Db CGCTGCCGAGTTCTTCAACAACTTTTCTGATCACCACATGACGAGTGTGAGAAAG 1323
Qy 1094 TGATGACACGACACAAACCATCCAGTGTGATCATGTGAGTGTGGCAACGAC 1153
Db TGGTCGTGAGGACAAAGACACCCCGCGTGTGATGTGTCTGTGGCCAAAGAGCTTG 1383
Qy 1154 AGTCCAACCATCCAGACCGCGAGGGTTTCTTCAAAAGCCCTTATGAGACTGGCAATGAAA 1213
Db GTTCCCACTAGAACTCTGCTGCTACTTGAAGTGTGATTCGCTACACCAATCTCT 1443
Qy 1214 TGGATCGAACACGCCCCGTTGTTCATGTGAGATGATGAGACCGACACGACGAGAACAA 1273
Db TGGACCCCTCCGCGCTGTGACCTTTGTGAGCACTTAACATGAG----- 1491
Qy 1274 GAGACGTGGCGCTGAAGTACTTGACATCGTGTGTGAACAGGTACTACGGCTGTACA 1333
Db CAGACAAAGGGGCTCCGTATGTGTGATGTGATCTGTTGAACAGCTACTCTCTGTGATC 1551
Qy 1334 TCTATCAGGAGAGATAGAAAGAGCTTCAGAGCTTGGAAAAACATAGAAAGCTCT 1393

Db 1552 ACAGTACGGGACCTGAGTTGATTTCAGCTGACGCTGGCCACCCAGTTTGAAACTGGT 1611
 Qy 1394 ATGCAAGGACAGAAAGCCCATCTTTGTGCAGAAATTCGTCGAGCGATAGTGGCA 1453
 Db 1612 ATAAAGATATCAGAAAGCCCATTTATTCAGAGCAGATAGAGCAAGAAACATTTGACGGGT 1671
 Qy 1454 TCCACTACGATCCACCTCAAAATGTTCTCCAGAGATACCAAGAGAGCTGTGAAAG- 1512
 Db 1672 TTACACGAGATCCACCTGTATGTTCTAGAGATGACGAAAGATGCTGTAGACAGT 1731
 Qy 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGGAAACACCGTGTGGGCT 1567
 Db 1732 ACCATCTGGGCTCTGATCAAAACGACGAAATATGTGTGAGAGCTCATTTGAAAT 1791
 Qy 1568 TTGCAATTTTAAAGCTCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGGCTGTTT 1627
 Db 1792 TTGCGCATTTCAATGACTGAAACAGTCAACGAGAGAGTCTGGGGAATTAAGGGGATCT 1851
 Qy 1628 TCACAAAGACAGACAAACCAACTGCTGTGCTCATGTAGAGA 1671
 Db 1852 TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTGCGA 1895

RESULT 12
 ADP10326
 ID ADP10326 standard; DNA, 2191 BP.
 AC ADP10326;
 XX
 DT 12-AUG-2004 (first entry)
 DE
 XX Reference mRNA sequences for marker probe #3.
 DE
 XX transplant rejection; immune system; rheumatoid arthritis; lupus;
 KW inflammatory bowel disease; multiple sclerosis; HIV; AIDS; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO2004042346-A2.
 XX
 PD 21-MAY-2004.
 XX
 XX 24-APR-2003; 2003WO-US012946.
 XX
 XX 24-APR-2002; 2002US-0031831.
 PR 20-DEC-2002; 2002US-00325899.
 XX
 XX (EXPR-) EXPRESSION DIAGNOSTICS INC.
 PA
 P1 Wohlgemuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;
 P1 Rosenberg S;
 XX
 DR WPI, 2004-400724/37.
 XX
 PT Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
 PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
 PT rejection, in an individual, comprises detecting the expression level of
 PT the genes.
 XX
 XX Claim 80; SEQ ID NO 335; 1762bp; English.

CC viral, bacterial or fungal infection. The present sequence represents a
 CC mRNA reference sequence for a 50 mer oligonucleotide marker for diagnosis
 CC and monitoring of allograft rejection and other disorders.
 XX
 SQ Sequence 2191 BP, 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;
 Query Match 10.2%; Score 172.4; DB 12; Length 2191;
 Best Local Similarity 52.1%; Pred. No. 8e-43;
 Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;
 Qy 737 TGGAACTTGAAGAAAGACGATGACCTGTGACATCGAATCGAATGAAATGACCTGGAGC 796
 Db 964 TGGGGCTGTGTCTGATCTTACACATCCCTGTGGGATCGCATGTGTGTGACCA 1023
 Qy 797 AGAAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAGACGAG 856
 Db 1024 AGAGCAAGTTCCTCATCAATGGGAAACCTTCTATTTCCACGCTGTCAACAGCATGAG 1083
 Qy 857 AATTCCTCGTCTGGGGGACGACCTTTTATTCATTGATGATTAAGACTTCAACCTTC 916
 Db 1084 ATGCGGACATCCGAGGGAAGGCTTCACTGGCCGCTGTGTGAAGACTTCAACCTTC 1143
 Qy 917 TGAAGTGAATCAACGGAATTTCTTCAAGACCTTCACTATCTTACATGAAAGTGGC 976
 Db 1144 TTGCTGCTGTGTGTGCAACGCTTTCCTTCCATCCAGCCACTACCTTATGCAAGAGTGA 1203
 Qy 977 TGGATTTGCGCAGACGACTCGGAATCCTTGTGATGAGAGACCCGCACTTGGATCA 1036
 Db 1204 TGCAGATGTGTAACGCTATGAGATTTGTGTATGATGATGATGATGATGATGATGATG 1263
 Qy 1037 CAAGTACCA---CTTACATCCGAGACTCAGAAATGACAGAAAGCAACATTAAGAA 1093
 Db 1264 CGCTGCCGAGTCTTTCACAAAGTTTCTCTGATCACCACAGAGAGTGAAGAAAG 1323
 Qy 1094 TGAATGACAGACAAAGAACATCCCATGTGTATCATGTGAGTGTGGGAGCAACCA 1153
 Db 1324 TGGTGTGAGGAGCAAGAAACCAACCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1383
 Qy 1154 AGTCCAAACCATCCAGACGGGAGGGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAA 1213
 Db 1384 CGTCCACCTAGAAATGTGCTGCTACTTATGAAAGTGTGATCGTCACACCAATCCT 1443
 Qy 1214 TGGATGAAACAGCCCGCTGTGTATGATGAGACATGATGAGACGACAGAGAAACAA 1273
 Db 1444 TGGACCCCTCCGCGCTGTGACTTGTGAGCAACTTAATATGACG----- 1491
 Qy 1274 GAGAGCTGGCGCTGAAGTACTTGCATATGCTGTGTGTGAACGCTACTACGCTGTGACA 1333
 Db 1492 CAGACAAAGGGGGCTCCGATGTGTGATGTATCTGTTGAACAGCTACTACTTGTGTATC 1551
 Qy 1334 TCTATCAGGGAAGGATAGAAAGGACTTCAAGCTCTGAAAAAAGACATAGAAAGCTCT 1393
 Db 1552 ACAGTACGGGACCTGAGATTGATTCAGCTGACGTGGCCACCCAGTTTGAACATGGT 1611
 Qy 1394 ATGCAAGGACAGAAAGCCCATCTTTGTGCAGAAATTCGTCGAGCGATAGTGGCA 1453
 Db 1612 ATAAAGATATCAGAAAGCCCATTTATTCAGAGCAGATAGAGCAAGAAACATTTGACGGGT 1671
 Qy 1454 TCCACTACGATCCACCTCAAAATGTTCTCCAGAGATACCAAGAGAGCTGTGAAAG- 1512
 Db 1672 TTACACGAGATCCACCTGTATGTTCTAGAGATGACGAAAGATGCTGTAGACAGT 1731
 Qy 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGGAAACACCGTGTGGGCT 1567
 Db 1732 ACCATCTGGGCTCTGATCAAAACGACGAAATATGTGTGAGAGCTCATTTGAAAT 1791
 Qy 1568 TTGCAATTTTAAAGCTCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGGCTGTTT 1627
 Db 1792 TTGCGCATTTCAATGACTGAAACAGTCAACGAGAGAGTCTGGGGAATTAAGGGGATCT 1851
 Qy 1628 TCACAAAGACAGACAAACCAACTGCTGTGCTCATGTAGAGA 1671
 Db 1852 TCACCTGGCAGAGACAAACCAAAAGTGACGCTTCTTTGCGA 1895

RESULT 13

AAFS7574
ID AAF57574 standard; DNA; 3525 BP.

AAFS7574;

11-JUN-2001 (first entry)

Mutant heavy chain construct 431 nucleotide sequence.

Fusion protein; cell surface antigen; cancer; monoclonal antibody;
heavy chain construct; carcinoembryonic; mutant; de.

Mammalia.

WO200119842-A1.

22-MAR-2001.

18-SEP-2000; 2000WO-US025558.

17-SEP-1999; 99US-00399079.

(GEN2) GENZYME TRANSGENICS CORP.

Pollock D, Meade HM, Bossett K;

WPI; 2001-257871/26.

P-PsDB; AAB62273, AAB62274, AAB62275, AAB62276.

New fusion protein, useful for killing diseased cells e.g. cancer cells,
comprises fused first and second units, such that the protein assembles
into complex which optimizes activity of multimeric form of second unit.

Example 5; Fig 5; 88pp; English.

The invention relates to a fusion protein (I) comprising a first member
fused to a second member, where the first and second members are chosen
such that (I) assembles into a complex having a number of subunits which
optimizes activity of the multimeric form of the second member. (I) is
useful for selectively killing an aberrant or diseased cell which
expresses a target antigen on its surface, e.g., a cancer cell expressing
a cell surface antigen. (I) is also useful for detecting in vitro or in
vivo the presence of target antigen in a sample, e.g., for diagnosing a
disease, by contacting a sample or a control sample that allows
interaction of (I) which is labeled, and detecting formation of a
complex. (I) is also useful for selectively directing (e.g., localizing)
the second unit of (I) to the vicinity of an undesirable cell. The
present sequence represents a nucleotide sequence for the mutant heavy
chain of humanised anti-carcinoembryonic antigen antibody 431 lacking the
hinge region

Sequence 3525 BP; 763 A; 1031 C; 981 G; 730 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 4; Length 3525;

Best Local Similarity 52.1%; Pred. No. 1.1e-42;

Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGAAGCTGGAAGAAAGAGTACCTGAGCATGCGAATGAGACGATGACGCTGGAGC 796
DB 2500 TGGGCGCTGTGTCTACTTCTTACACATCCCTGTGGAGATCGCATGTGGCTGCACCA 2559
QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAAGCAGAG 856
DB 2560 AGAGCAGATCTCTATCATGAGGAAACCTTTCTATTCCAGGATCAACAAGATGAGG 2619
QY 857 AATTCCCGCTTGGGGCAGGGCAGCTTTTATCATTTGATGATTAAGAAGCTTCACTTC 916
DB 2620 ATGGGAGATCTCGAAGGAGGCTTTCGACTGGCCGCTGTGTGAAGAGCTTCAACCTGC 2679
QY 917 TGAAGTGATCAAGCGAATCTTTCAAGAGCTCTCACTATCTTACAGTGAAGGTGGC 976

DB 2680 TTGCTGTGCTGTGTGTGTCACACGCTTTCCGTACAGCAGCTACCTTATGACAGGAGATGA 2739
QY 977 TGGATCTTCCGACAGACTCGGAATCTTTGTGATGACGAAGCCCCGACGTTGTATCA 1036
DB 2740 TGCAGATGTGTGACCGCTATGTGGATGTGTGTATGATGATGATGATCCCGCGTGGCCTGG 2799
QY 1037 CAAGGTACCA---CTACAAATCCCGAGACTCCGAAATATGACGAAGACCAACTAAGAA 1093
DB 2800 CGCTCCGAGTCTTTCACCAACGTTTCTGTGATCACACATGCAAGGTGATGAAAG 2859
QY 1094 TGAATGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
DB 2860 TGTGTGATGAGGACAAAGAACCAACCCCGGCTGTGTATGTGTGTGTGTGCAAGGCTTG 2919
QY 1154 AGTCCAACTTCAGACCGCGAGGGTCTTTCAAAGCCCTTTATGAGACTGCGAATGAA 1213
DB 2920 CGTCCCACTTGAATCTGTGCTGTACTTGTGAATGATGTGATCGCTCACACCAATGCT 2979
QY 1214 TGGATGAAACAGCCCGCTGTGTGATGTGATGATGATGATGATGATGATGATGATGAT 1273
DB 2980 TGAACCTCTCCCGGCTGTGACTTTGTGAGCACTTAAGTATGAG----- 3027
QY 1274 GAGAGTGGCGCTGAAAGTACTTGCATCTGTGTGTAACAGGACTTACGCTGTATCA 1333
DB 3028 CAGACAAAGGGGCTCCGATGTGTGATGTGTGATGTGTGATGTGTGATGTGTGATGTG 3087
QY 1334 TCTATCAGGAAAGATAGAAAGAGACTTCAAGCTCTGGAAGAAACATAGAAAGCTCT 1393
DB 3088 ACGACTACGGGACACTGAGTGTATTCAGCTGACAGCTGGCCACCAAGTTTGAAGCTGGT 3147
QY 1394 ATGCAAGCAGAAAGAGCCCATCTTTGACAGAAATTTGGTGGGAGCGGATAGTGGCA 1453
DB 3148 ATAGAAATATACAGAAAGCCCATTTTACAGACCGGATAGAGCAGAAACGATTTGAGGGT 3207
QY 1454 TCACATGATCCAGCTCAAAATGTTCTCCGAAAGATCAAGCAGAGCTCGTTGAAAG- 1512
DB 3208 TTCACAGAGATCCACTTGTATGTTCACACTGAAAGATCAAGAAAGTTGCTAGAGCAGT 3267
QY 1513 ACGATCAGGCTCTTTTGAAGAA---AGACTACATATCGAAGACACAGCTGTGGGCT 1567
DB 3268 ACGATCTGGTCTGGATCAAAAGACGAAATATGTGTTGGAGAGCTCATTTTGAATTT 3327
QY 1568 TTGCAATTTTAAAGACTCTGAAATGTGAAAGAACCATTTCTAACACAAAGGTTT 1627
DB 3328 TTGCGAATTTATGATGATCAAGTCAACGACGAGAGTGTGCTGGGAAATAAAGGGGATCT 3387
QY 1628 TCACAAGAGACAGCAACCAACCTCGTTGCTCATGTACTGAGA 1671
DB 3388 TCACTCGGAGAGACCAACCAAAAGTGCAGCGCTTCTTTTCCGA 3431

RESULT 14

AAFS7573
ID AAF57573 standard; DNA; 7328 BP.

AAFS7573;

11-JUN-2001 (first entry)

Heavy chain construct 431a nucleotide sequence.

Fusion protein; cell surface antigen; cancer; monoclonal antibody;

heavy chain construct; carcinoembryonic; de.

Mammalia.

WO200119842-A1.

22-MAR-2001.

18-SEP-2000; 2000WO-US025558.

XX

XX Gehrman M, Seemann G, Bosslet K, Czech J;
 DR WPI, 1994-111012/14.
 DR P-PSDB; AARS0092.
 XX New fusion protein contg. enzyme for prodng activation - coupled to
 PT antigen binding component, esp. sfv antibody fragment, partic. for
 PT treatment of tumours.
 XX Claim 16; Page 12-15; 35p; German.
 XX The sequence AAO58896 codes for a humanised sfv-fragment against CEA and
 CC a human beta-glucuronidase. The resulting fusion protein is useful for
 CC targeting beta-glucuronidase to cancer cells expressing CEA, where the
 CC enzyme is able to convert a prodng into its active form. Any fusion
 CC protein not bound to tumour can be removed by internalisation via the
 CC mannose-6-phosphate and galactose receptors. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 XX
 SQ Sequence 3314 BP; 789 A; 924 C; 890 G; 711 T; 0 U; 0 Other;

Query Match 10.1%; Score 170.8; DB 2; Length 3314;
 Best Local Similarity 52.0%; Pred. No. 3.3e-42;
 Matches 491; Conservative 0; Mismatches 432; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGAAAGAGTACCTCTGAGATCGAATCAAGATGATGAGGAG 796
 DB 2146 TGGGCGCTGTCTCTCACTTCTACACATCCCTGTGGGATCGCATGTGCTGCACCA 2205

QY 797 AGAAGAGCTCTATCTGAACCGGAAACCTGTCTTTTGAAGGCTTTGGAAGCAGAG 856
 DB 2206 AGAGCAGTTCTCTATCAATGAGAAACCTTTCTATTTCCAGGTGTCACAGCATGAGG 2265

QY 857 AATTCCTCGTCTGAGGAGGAGCACTTTTATCATTTGATGATTAAGACTTCAACTTC 916
 DB 2266 ATGCGGACATCCGAGGAAAGGCTTGAAGTGGCCCTGCTGTGAAGACTTCAACTTC 2325

QY 917 TGAAGTGAATCAACCGAATTTCTTCAAGACCTTCACTATCTTACAGTGAAGAGTGGC 976
 DB 2326 TTCCGTGCTTGTGTCACCAAGCTTTCGTAACAGCACATACCCATGAGAGAGTGA 2385

QY 977 TGGATCTTCCGAGCAGATCCGGAATCTTGTGATGAGAGAGCCCGCAGCTTGTGATCA 1036
 DB 2386 TGCAGATGTGTCGCTATGAGGATGTGTGATGATGATGATGATGATGATGATGATGAT 2445

QY 1037 CAAGGTACCA---CTACATCCGAGACTCAGAGATGAGAGAGAGAGAGAGAGAGAG 1093
 DB 2446 CGCTGCCGAGTTCTTCAACAGCTTCTCTGATCACCAGATGAGAGAGAGAGAGAG 2505

QY 1094 TGAATGACAGACAG 1153
 DB 2506 TGGTGTGAGGAG 2565

QY 1154 AGTCCACATTCACAG 1213
 DB 2566 GTTCCACATTCAG 2625

QY 1214 TGAATGACAG 1273
 DB 2626 TGGAGCCCTCCCGGCTGTGAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2673

QY 1274 GAGAGTGGCGCTGAAGTATCTTGAATCTGTGTGAACAGTACTTACGCTGTGATCA 1333
 DB 2674 CAGACAG 2733

QY 1334 TCTATCAGGAG 1393
 DB 2734 ACAGCTACGAG 2793

QY 1394 ATGCAAGGAG 1453
 DB 2794 ATGAG 2853

QY 1454 TCCACTAGATTCACCTCAATGTTTCTCCGAAGATACCAAGCAGAGCTCGTTGAAG- 1512
 DB 2854 TTACACAGAGATCCACTGTGATGTTCACTGAAGAGTACAGAGAAAGTGTGAGAGCAGT 2913

QY 1513 AGCATCAGGCTCCTTTGAAAAA-----AGACTACATCATGCGAAGACAGTGTGGGCT 1567
 DB 2914 ACCATCTGGTCTGATCAAAAAAGCAGAAATATGTGTGAGAGAGCTCATTTGAAATT 2973

QY 1568 TTGAGATTTTAAAGACTCCTCAGAAATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1627
 DB 2974 TTGCCGATTTTATGATGATGAACAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3033

QY 1628 TCACAG 1671
 DB 3034 TCATCGGAG 3077

Search completed: January 24, 2005, 03:56:03
 Job time : 866 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: January 24, 2005, 03:20:50 ; Search time 156 Seconds

(without alignments)
7695.657 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689
Sequence: 1 atggtaagaccgcaacgaacaa.....gaagactgtgagatgaggtc 1689

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	4	US-09-270-957-14
2	172.4	10.2	1956	4	US-09-715-858-1
3	172.4	10.2	2191	3	US-09-039-555B-12
4	159.6	9.4	1947	4	US-09-715-858-3
5	141.4	8.4	1887	3	US-09-149-727-7
6	130.4	7.7	1888	4	US-09-270-957-27
7	112	6.6	14683	2	US-08-819-866-1
8	112	6.6	14683	2	US-09-023-715-1
9	112	6.6	14683	4	US-09-343-485A-1
10	110.8	6.6	3035	2	US-08-723-624-18
11	110.8	6.6	3169	3	US-08-630-820-5
12	110.8	6.6	3169	4	US-09-273-453-5
13	110.8	6.6	3824	2	US-08-723-624-19
14	110.8	6.6	4652	4	US-09-893-525-36
15	110.8	6.6	5390	4	US-09-893-525-41
16	110.8	6.6	5418	4	US-09-893-525-38
17	110.8	6.6	11978	4	US-09-792-568-8
18	110.8	6.6	14348	4	US-09-792-568-9
19	110.8	6.6	18986	2	US-08-819-866-2
20	110.8	6.6	18986	2	US-09-023-715-2
21	110.8	6.6	18986	4	US-09-343-485A-2
22	110.8	6.6	32798	4	US-09-604-694B-1
23	109	6.5	1806	4	US-09-270-957-7
24	109	6.5	1854	3	US-09-149-727-3
25	109	6.5	2100	4	US-09-270-957-1
26	109	6.5	6029	3	US-09-149-727-1
27	104.4	6.2	8012	3	US-09-182-117-1

28	104.4	6.2	8012	4	US-09-434-039A-1	Sequence 1, Appli
29	104.4	6.2	8418	3	US-09-182-117-5	Sequence 5, Appli
30	104.4	6.2	8418	4	US-09-434-039A-5	Sequence 5, Appli
31	104.4	6.2	8798	3	US-09-182-117-4	Sequence 4, Appli
32	104.4	6.2	8798	4	US-09-434-039A-4	Sequence 4, Appli
33	98	5.8	2141	4	US-09-445-283C-27	Sequence 27, Appli
34	98	5.8	2633	1	US-08-452-267-2	Sequence 2, Appli
35	98	5.8	2633	3	US-09-123-644-2	Sequence 1, Appli
36	98	5.8	4544	4	US-09-488-270A-1	Sequence 21, Appli
37	98	5.8	4947	4	US-09-118-276-21	Sequence 21, Appli
38	98	5.8	5534	1	US-08-452-267-3	Sequence 3, Appli
39	98	5.8	5534	3	US-09-123-644-3	Sequence 3, Appli
40	98	5.8	5560	3	US-08-817-188-5	Sequence 5, Appli
41	98	5.8	5642	1	US-08-318-772A-2	Sequence 2, Appli
42	98	5.8	5897	4	US-09-097-319A-26	Sequence 26, Appli
43	98	5.8	5897	3	US-09-643-971-26	Sequence 26, Appli
44	98	5.8	6898	3	US-09-097-319A-27	Sequence 27, Appli
45	98	5.8	6898	4	US-09-643-971-27	Sequence 27, Appli

ALIGNMENTS

RESULT 1
US-09-270-957-14
Sequence 14, Application US/09270957
Patent No. 6641996
GENERAL INFORMATION:
APPLICANT: Richard A. Jefferson and Jorge E. Mayer
TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
FILE REFERENCE: 190106.405C1
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 112
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 1689
TYPE: DNA
ORGANISM: Thermotoga maritima
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(1689)
OTHER INFORMATION: n = A,T,C or G
US-09-270-957-14

Query Match	99.9%	Score 1688	DB 4	Length 1689
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1689	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	ATGGTAAGACCGCAACGAAGAGATTATTTCTTATCTTGAATGAGATTGGAAT	60	
DB	1	ATGGTAAGACCGCAACGAAGAGATTATTTCTTATCTTGAATGAGATTGGAAT	60	
QY	61	CTTGAAAGTACCGAAGAGACGACATCGCCCTCTCTGGAAGTGGAAAGAGACATAC	120	
DB	61	CTTGAAAGTACCGAAGAGACGACATCGCCCTCTCTGGAAGTGGAAAGAGACATAC	120	
QY	121	CAGATCTGTGCTAGGAAGAGACCCCTTCACTTCAAAAACCACTTACGTTCCGAG	180	
DB	121	CAGATCTGTGCTAGGAAGAGACCCCTTCACTTCAAAAACCACTTACGTTCCGAG	180	
QY	181	NAACCTTCAAAAACATCAGACTTACTTGTGCGTGAACACGAGTCGAGGTC	240	
DB	181	NAACCTTCAAAAACATCAGACTTACTTGTGCGTGAACACGAGTCGAGGTC	240	
QY	241	TTTCTCAACGGAAGAAAGTGGAGAAATACATTTGATCTTCCCTTGAAGTAAAT	300	
DB	241	TTTCTCAACGGAAGAAAGTGGAGAAATACATTTGATCTTCCCTTGAAGTAAAT	300	
QY	301	GTGACGGGGAAGTGAATTCGAGAGAGAACGACTCAGGGTGTGTTGAGAACGATTG	360	
DB	301	GTGACGGGGAAGTGAATTCGAGAGAGAACGACTCAGGGTGTGTTGAGAACGATTG	360	

```

Qy 361 AAAGTGGAGAGATTTCCCTCGAAGTTCCAGACAGCCGCACTCACACCGTGGGATTTTT 420
Db 361 AAAGTGGAGAGATTTCCCTCGAAGTTCCAGACAGCCGCACTCACACCGTGGGATTTTT 420
Qy 421 GGAAGTTTTCACCTGCAAACTTGCATCTTCCCTCACTGCTGGAATCATTAAGGCTGTT 480
Db 421 GGAAGTTTTCACCTGCAAACTTGCATCTTCCCTCACTGCTGGAATCATTAAGGCTGTT 480
Qy 481 CTGATAGATTTCACAGACCAAGGAGATCTGACATCTGGGTGACACAGATGAGTCT 540
Db 481 CTGATAGATTTCACAGACCAAGGAGATCTGACATCTGGGTGACACAGATGAGTCT 540
Qy 541 GAACCGGAGAGAACTTGGAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 600
Db 541 GAACCGGAGAGAACTTGGAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 600
Qy 601 GGACGAGAGATGACGATCAAACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 GGACGAGAGATGACGATCAAACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy 661 TTGCTCGAAGGGGAGTTTATCTCGAAGAACCCAGGTTCTGGAGCTTCGAAATCATAT 720
Db 661 TTGCTCGAAGGGGAGTTTATCTCGAAGAACCCAGGTTCTGGAGCTTCGAAATCATAT 720
Qy 721 CTTTATCCTCTCAAGGTGGAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 721 CTTTATCCTCTCAAGGTGGAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Qy 781 ACAGTACAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Db 781 ACAGTACAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
Qy 841 TTTGGAAAAGCAGAGAGATTTCCCTGTTGGGGGAGGGGACCTTTTATCCATGATGATA 900
Db 841 TTTGGAAAAGCAGAGAGATTTCCCTGTTGGGGGAGGGGACCTTTTATCCATGATGATA 900
Qy 901 AAAGACTTCAACTTCTGAAGTGGATCAACGCGAATCTTTAGAGACCTTCATATCT 960
Db 901 AAAGACTTCAACTTCTGAAGTGGATCAACGCGAATCTTTAGAGACCTTCATATCT 960
Qy 961 TACAGTGAAGAGTGGTGGATCTTGGCGACAGACTCGGAATCTTGTGATGAGAGAGCC 1020
Db 961 TACAGTGAAGAGTGGTGGATCTTGGCGACAGACTCGGAATCTTGTGATGAGAGAGCC 1020
Qy 1021 CCGGACCTTGGTATCAAGATGACCACTAATATCCGAGACTCGAAGATGAGAGAGAG 1080
Db 1021 CCGGACCTTGGTATCAAGATGACCACTAATATCCGAGACTCGAAGATGAGAGAGAG 1080
Qy 1081 AACATAAGAAAGATGATGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 AACATAAGAAAGATGATGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy 1141 GCGAAGCAACAGAGTCCAACTCAGACGCGGAGGGGTTCTTCAAGCCCTTTATGAG 1200
Db 1141 GCGAAGCAACAGAGTCCAACTCAGACGCGGAGGGGTTCTTCAAGCCCTTTATGAG 1200
Qy 1201 ACTGCCAATGAATGATGCAAGACGCGCGGTTCTCATGTGTGAGCAATGATGAGAGCC 1260
Db 1201 ACTGCCAATGAATGATGCAAGACGCGCGGTTCTCATGTGTGAGCAATGATGAGAGCC 1260
Qy 1261 GACGAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 GACGAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy 1321 TACGGCTGGTATCATCTATCAAGGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 TACGGCTGGTATCATCTATCAAGGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy 1381 ATGAAGAGAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 ATGAAGAGAGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

```

```

Qy 1441 GCGATAGCTGGATCCACTTACATCCACTCAATGTTCTCCGAGAGAGTACCAAGCAGAG 1500
Db 1441 GCGATAGCTGGATCCACTTACATCCACTCAATGTTCTCCGAGAGAGTACCAAGCAGAG 1500
Qy 1501 CTCGTTGAAAAGAGAGATCAGAGGCTCTTTGAAAAAGAGACTATCATCGAAGACAGCTG 1560
Db 1501 CTCGTTGAAAAGAGAGATCAGAGGCTCTTTGAAAAAGAGACTATCATCGAAGACAGCTG 1560
Qy 1561 TGGGCTTTGACAGATTTTAAAGTCTCTGAGATGTGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1561 TGGGCTTTGACAGATTTTAAAGTCTCTGAGATGTGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy 1621 GGTGTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Db 1621 GGTGTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy 1681 AGTGAGGTT 1689
Db 1681 AGTGAGGTT 1689

RESULT 2
US-09-715-858-1
; Sequence 1, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VARIANTS FOR THE
; TITRE OF INVENTION: TREATMENT OF LYSSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/09/715, 858
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patencin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1956)
US-09-715-858-1

Query Match 10.2%; Score 172.4; DB 4; Length 1956;
Best Local Similarity 52.1%; Pred. No. 2.1e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

Qy 737 TGGAACTTGAAGAAAGACAGATCACTTGGAGATCGGAATCAGAAAGATCAGCTGGAGC 796
Db 737 TGGAGGCTGTGTCTGATCTTCACTCAACACTCCGTGGGGATCCGCACTGTGGCTGACCA 997
Qy 797 AGAAGAGGCTCTATCTGAAGCGGAAACCTGTCTTTTGAAGGGCTTTGAAAACAGAGAG 856
Db 797 AGAGCAAGTTCCTCATCAATGAGAAACCTTTGATTTTCAAGGCTGTCAACAGAGAGAG 1057
Qy 857 AATTCCTGCTTGGGGAGGGGCACTTTTATCATATGATGATGATGATGATGATGATGATG 916
Db 857 ATCCGAGATCCAGAGGAGAGGGGCTTGAAGCTGGCGCTGTGGTGAAGAGAGAGAGAGAG 1117
Qy 917 TGAAGTGAATCAAGCGGAGATCTTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 976
Db 917 TGGCTGGCTGTGTGCAAGCTTTCGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
Qy 977 TGGATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1036
Db 977 TGAAGATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237
Qy 1037 CAAAGTACCA---CTTCAATCCGAGAGCTCAGAGATGACAGAGAGAGAGAGAGAGAGAG 1093
Db 1037 CCGTGGCGAGTCTTCAACAGGTTTCTGTGATCAGACATGAGAGAGAGAGAGAGAGAGAG 1297

```

```

Qy 1094 TGATCGACAGACACAAAGACCATCCCACTGATCATGTGAGTGTGCGCAACGACG 1153
Db 1298 TGGTGGCGTAGGAGACAAAGACCAACCCCGCGTGTGATGTGTGTGGCCAAACGAGCTG 1357
Qy 1154 AGTCCAAACCATTCAGACCGGAGGGTTTCTTCAAGCCCTTATAGACTGCGCAATGAAA 1213
Db 1358 CGTCCCACTTAGAATCTGTGTGCTACTACTTGAATGATGTATGCTACACCAATCTCT 1417
Qy 1214 TGGATCGAACACGCCCCCTTGTGATGTGAGCATGAGACGACACGACGAGAACAA 1273
Db 1418 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCACTTAATATGACAG----- 1465
Qy 1274 GAGAGTGGCGCTGAAGTACTTCCAGATCGTGTGTGTAACAGTACTACGCTGTACA 1333
Db 1466 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTTGTGTATC 1525
Qy 1334 TCTATCAGGAAAGATAGAAAGAGACTTCAAGCTCTGGAATAAAGACATAGAGAGCTCT 1393
Db 1526 ACGACTACGGGCACTGAGAGTTGATTCAGCTGACGTCGACCCAGTTTGAAGACTGCT 1585
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTTGTGACAGAAATTCGTCGCGACGCGATAGCTGCA 1453
Db 1586 ATAAGAAATATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAAGATTCAGGGGT 1645
Qy 1454 TCCACTAGATTCACACTCAATGTCTCGAAGATACCAAGACAGAGCTCGTTGAAAAG- 1512
Db 1646 TTCAACAGAGATTCACACTCTGATGTCTACTGAAAGATACAGAAAGTCTGTAGAGCAGT 1705
Qy 1513 ACGATCAGGCTCTCTTTGAAAAA-----AGACTACATCATCGGAACACAGGTGTGGGCT 1567
Db 1706 ACGATCTGGTCTGTGATCAAAAAGCAGAAATAATGTGTGAGAGCTCATTTGGAATT 1765
Qy 1568 TTGAGATTTTAAAGCTCTCTCAGAAATGTGAGAAAGCCATTTCTCAACCAAGAGGTGTTT 1627
Db 1766 TTGCGGATTTATGACTGAAACGTCACCGACGAGAGTGTCTGGGGAATTAAGAGGAGTCT 1825
Qy 1628 TCACAAAGACAGACAAACCAACCTGTTGCTCATGTACTGAGA 1671
Db 1826 TCACCTCGGACAGACAAACCAAAAGTGCAGCGTTCCTTTTGGA 1869

```

RESULT 3

US-09-039-555B-12

Sequence 12, Application US/09039555B

Patent No. 6033856

GENERAL INFORMATION:

APPLICANT: Koerner, Kathrin

APPLICANT: Mueller, Rolf

APPLICANT: Sadleir, Hans-Harald

TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/039.555B

FILING DATE: 16-MAR-1998

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE 19710643.9

FILING DATE: 14-MAR-1997

ATTORNEY/AGENT INFORMATION:

```

NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 016779/0131
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2191 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-039-555B-12

Query Match      10.2%; Score 172.4; DB 3; Length 2191;
Best Local Similarity 52.1%; Pred. No. 2.3e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

Qy 737 TGAACCTTGAAGAAAGACGATACACTGTGACATGGAATGAAAGATCAGCTGGAGCG 796
Db 964 TGGGGCTGTGTGATCTTGAACACTCCCTGTGGGATCCGACTGTGCTGACCA 1023
Qy 797 AGAAGGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGCAGAG 856
Db 1024 AGAGCACTTCTCATCATGAGGAAACCTTCTTATTTCCAGGTGTCAACAGCATGAGG 1083
Qy 857 AATTCCTCCGTTCTGGGGGAGGGGACCTTTTATCCATGATGATTAAGACTTCAACTTC 916
Db 1084 ATGCGGACATCGAGAGGAGGAGGCTTGACAGCTGGCGCTCTGTGAAGACTTCAACTTC 1143
Qy 917 TGAAGTGATCAAGCGAATCTTTGAGAGCTCTGACTATCTTTACAGTGAAGAGTGC 976
Db 1144 TTCCGTGGCTGTGTCGCAACCTTTCCGTACAGCCACTACCCCTATGACAGAGAAATGA 1203
Qy 977 TGGATCTTGGCAGACAGATCGGAATCCTGTGTGATGAAGAAAGCCCGACCTGTGTATCA 1036
Db 1204 TGCAGATGTGTACCGCTATGAGATTTGTGTATGTATGATGATGTCCCGCGGTGCTGG 1263
Qy 1037 CAAGTACCA---CTACATCCCGAGACTCAGAAAGATGACAGAAAGCAATTAAGAA 1093
Db 1264 CGCTCCGCAATTTCTTCAACAACGTTTCTCTGCAATCACAATGAGATGATGAAGAA 1323
Qy 1094 TGATCGACAGACACAAAGACCATCCAGTGTATCATGTGAGTGTGCGAAGCAAG 1153
Db 1324 TGGTGGCGTAGGAGACAAAGACCAACCCCGCGTGTGATGTGTGTGCTGCAAGAGCTG 1383
Qy 1154 AGTCCAAACATCCAGACCGGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA 1213
Db 1384 CGTCCCACTTAGAATCTGTGCTGCTACTACTTGAAGATGTATGCTGCACACCAATCT 1443
Qy 1214 TGGATCGAACACGCCCCGTTGTGATGTGAGCATGATGAGACGACACGACGAGAACAA 1273
Db 1444 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCACTTAATATGACAG----- 1491
Qy 1274 GAGACGTGGCGCTGAAGTACTTGAACATCGTCTGTGTGAACAGGTACTACGCTGTACA 1333
Db 1492 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTGTAACAGCTACTACTTGTGTATC 1551
Qy 1334 TCTATCAGGAAAGATAGAAAGAGACTTCAAGCTCTTGAAAAAGACATAAGAGCTCT 1393
Db 1552 ACGACTACGGGACACTGAGATGTGATTCAGCTGACAGCTGGCACCAAGTTTGAAGACTGT 1611
Qy 1394 ATGCAAGGACAGAAAGCCCATCTTTTGTGACAGAAATTCGTCGAGACGCGATAGCTGCA 1453
Db 1612 ATAAGAAATATCAGAAAGCCCATTTATTCAGACGAGTATGAGCAGAAAGATTCAGAGGT 1671
Qy 1454 TCCACTAGATTCACACTTCAATGTCTTCCAGAGATCAAGACAGAGCTCGTTGAAAAG- 1512
Db 1672 TTCAACAGAGATTCACACTGTGATGTTCACCTGAAGATACAGAAAGTCTGTAGAGCAGT 1731
Qy 1513 ACGATCAGGCTCTTTTGAAGAA-----AGACTACATCATGTGGAACACACAGTGTGGGCT 1567

```

Db 1732 ACCATCTGGCTGATCAAAAACGAGAAATATGTGTTGAGAGCTCATTTTGAATT 1791
 Qy 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGAGGTGTTT 1627
 Db 1792 TTGGCGATTTTCACTGAAAGTCAACGAGAGAGTGTGGGAAATAAAAAGGGATCT 1851
 Qy 1628 TCACAGAGACAGACAAACCAAACTGTTGCTCATGTACTGAGA 1671
 Db 1852 TCACTGGAGAGACAAACCAAAAGTGAGCGTTCCTTTTGCA 1895

RESULT 4 US-09-715-858-3

; Sequence 3, Application US/09715858
 ; Patent No. 6582652
 ; GENERAL INFORMATION:
 ; APPLICANT: Podaskoff, Gregory
 ; APPLICANT: Watson, Gordon
 ; APPLICANT: Couro, Linda B.
 ; APPLICANT: Yang, Bin
 ; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
 ; TITLE OF INVENTION: TREATMENT OF LYSSOMAL DISORDERS
 ; FILE REFERENCE: 0800-0021
 ; CURRENT APPLICATION NUMBER: US/09/715,858
 ; CURRENT FILING DATE: 2000-11-14
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1947
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1947)
 ; US-09-715-858-3

Query Match 9.4%; Score 159.6; DB 4; Length 1947;
 Best Local Similarity 50.6%; Pred. No. 2.2e-39;
 Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

Qy 739 GAATTTGAAAAGAGAGTACCTCTGACATCGGAATCAGAACGATCAGCTGGACGAG 798
 Db 928 GAGTCTGTGACTGACTACACCTTCTTATCGGAGATTCAGACAGTGTGTCAAAAG 987
 Qy 799 AAGAGGCTCTATGACGAGAACTGTCTTTTGAAGGCTTTGAAAGCAGAGAA 858
 Db 988 AGCAGTTTCTCATTAAGGAGAGCCCTTCTATTTTCAAGGGTCAATTAAGCAGAGAT 1047
 Qy 859 TTCCCGCTTGGGAGAGGACCTTTTATCCATGTATGATTAAGACTTCAACCTTCTG 918
 Db 1048 TCAGATATCCGAGGAGAAAGCTTCACTGCGCTGCTGGTAAAGATTCAACCTGCTC 1107
 Qy 919 AAGTGTATCAAGGAGATTTCTTCAAGACCTCTCATATCTTTTCAAGTGAAGTGGCTG 978
 Db 1108 GTTGGCTCGGGGAAATTCCTTTCGTACAGCACCATCTTCACTCAGAGAGTACTT 1167
 Qy 979 GATCTTCCGACAGACTCGAATCTTGTGTATAGACGAGCCCGACGTTGG---TATC 1035
 Db 1168 CAGTCTGTATGACGATACGGGATTTGTGTATGTATGTATGTGTGTGTGTGTGTGTGT 1227
 Qy 1036 ACAAGTATCACTAATCCCGAGACTCAGAAATAGCAGAGCAACATMAAGAAATG 1095
 Db 1228 CTACTCGAGTTTGGCAACGAGTCACTTCGACACCACTAAGGTGTATGAGAGAGCTG 1287
 Qy 1096 ATGACAGACAGACAGAACTCCAGTGTATCATGTGTGTGTGTGTGTGTGTGTGTGTGT 1155
 Db 1288 GTTGGCGGAGCAAAATACCCCTGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1347
 Qy 1156 TCACACCATCAGACGCGGAGAGGTTTCTTCAAGCCCTTTATGAGATGCGCAATGAATG 1215
 Db 1348 TGTGCTTGAAACCCCGCGATTAATTATTAAGCGTGTATCAACCAACCAAGCCCTG 1407

Qy 1216 GATGAAACAGCCCGCTTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1275
 Db 1408 GACCTACCCGCTCCGTTGACCTTTTGTGAGCA-----ACGCAAAATGTATGCA 1455
 Qy 1276 GACGTGGCGTGAAGTCTTTCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1335
 Db 1456 GACCTGGGGGCCCCGTACGTAATGTATCTGTGTAAACGTAATCTTTTCTGTGTATCAT 1515
 Qy 1336 TATCAGGGAAGATAGAAAGAGACTTCAAGTCTGTGAAAACATATAGAAAGCTCTAT 1395
 Db 1516 GACTATGGGATTTGAGAGTATTCAGCCACACTGAATAGCCAGTTTGAACTGTAT 1575
 Qy 1396 GCAAGGCAAGAAAGCCCATCTTGTGTACAAATTCGTGCGAGCGGATAGTGCATC 1455
 Db 1576 AAGAGCATCAGAAAGCCGATTAATCCAGAGGATATGAGCAGACCAATCCAGGATC 1635
 Qy 1456 CACTAGATCCACTCAAAATGTTCTCCGAAGTACCAAGCAGTCTGTTGAAAAGAG 1515
 Db 1636 CAGAGAGCCGCTCGCATGTTCAGTGTGAGAGTACCAAGGCTGTCTGGAAATTAC 1695
 Qy 1516 ATCAGGCTCCTTTTGAAA-----AAAGACTATCATCGAACAACAGTGTGGCTTT 1569
 Db 1696 CATTAGTCTGTATCAGAAAGTAAAGATAGTGTGTGGAGAGCTCATCTGAAATTTC 1755
 Qy 1570 GCAGATTTTAAGCTCTCAGAAATGTGAGAAAGCCATTCTCAACCAAGGCTGTTTC 1629
 Db 1756 GCCGACTTATGACGAACCAAGTCCGCTGAGAGTATGGAACCAAGAGGAGATCTTC 1815
 Qy 1630 ACAAGAGACAGACCAACCACTGTTGCTCATGTACTGAGA 1671
 Db 1816 ACTGGCAGAGACAGCCCAAACTTGGCTTATTTTTCGA 1857

RESULT 5 US-09-149-727-7

; Sequence 7, Application US/09149727
 ; Patent No. 6391547
 ; GENERAL INFORMATION:
 ; APPLICANT: Jefferson, Richard A.
 ; APPLICANT: Kilian, Andrej
 ; APPLICANT: Keese, Paul Konrad
 ; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
 ; FILE REFERENCE: 190106.405
 ; CURRENT APPLICATION NUMBER: US/09/149,727
 ; CURRENT FILING DATE: 1998-09-08
 ; EARLIER APPLICATION NUMBER: US 60/058,263
 ; EARLIER FILING DATE: 1997-09-09
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 1887
 ; TYPE: DNA
 ; ORGANISM: Bacillus sp.
 ; US-09-149-727-7

Query Match 8.4%; Score 141.4; DB 3; Length 1887;
 Best Local Similarity 45.4%; Pred. No. 1.1e-33;
 Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

Qy 120 CCAGATCTGTGCTAGAAAGAGACCTTCACTAACAACCACTTCTAGTTCGGA 179
 Db 208 CAAAGAAATCCGCAACATATCGATATGTCTGTGCAAGGTGATCAGGCTCCGCG 267
 Qy 180 GNAATTTTCAAAAACATCAGCTTACTTGTGCTGCGGTGAACAGGACTGGAGT 239
 Db 268 CTATCTGAAGATCAGCTATCTGTGCTCCGCTCGCTCACTCAAAAGCAATGTT 327
 Qy 240 CTTCCTCAAGGAGAAAGGTGAGAGATCACTTGAATTAACCTTCTCTTGAAGTGA 299
 Db 328 CTATCTCAATGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 387
 Qy 300 TGTGACGGGAAAGTGAATCCGAGAGAAACAACTCAGGCTGTGTGTGTGAACAGATT 359

```

Db      388  AATCAACAACCTGCGTGGATGATGATGCGTCAACCGTCCGCGTGGACAATCT 447
Qy      360  GAAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGATTTT 419
Db      448  CGACATATAGCAACCTCCCGTGGGCTGTACAGCAGCGCCACGAAGAGGCGCTCGAAA 507
Qy      420  TGGAAATTTTCCACTGCAACTTGAATCTTCCCTAAGCGTGAATCATAGGCGCTGT 479
Db      508  AGTCAATGTGAACAAGCCGAACTTGACTTCTTCAATATGAGGCTTGCAACCGTCCGT 567
Qy      480  TCTGATAGATTCAAGACCAAGGATCTCAATCTGGGTGAGACAGAGTGAAGTC 539
Db      568  GAAATCTACAGCAGCCCGTTTACGTACGTGAGGACATCTCGTTGTGACCGACTTCA 627
Qy      540  TGAACCGGAGAAAGAACTTGGAAAAGTGAAGTGAAGTGAAGTCTCAAGAAAGCGT 599
Db      628  TGGCCCAACCGGAGCTGTGACCTATACGTGACTTTCAAGGCAAAAGCCGAGACCGTGA 687
Qy      600  GGGACAGGAGATGACGATCAAACTTGGAGAGAGAGAAAAGATTAGAACATCCACAG 659
Db      688  AGTGTGGTTCGTGATGAGAGAGGCAAGTGTCCGAGCACCGAGGCGTGAAGCGTGA 747
Qy      660  ATTCTCGAAGGGAGTTCACTCTCGAAAACGCGAGTTCTGAGCTCGAAGATCCATA 719
Db      748  CGTGAAGATTCGGAATGTCTATCTCTGGGAACCACTGAACAGTATCTTACCAATCA 807
Qy      720  TCTTTATCTCTCAAGTGGAACTTGAAMAAAGACGATACCTGTGACATCGAATCAG 779
Db      808  AGTGAACGTGGTGAACGAGCGACTGACATGATGTATGAAGACCGTTGGGCGT 867
Qy      780  AACGATCAGCTGGGAGAGAGAGGCTCTATCTGAACGGGAACTGTCTTTGAAGG 839
Db      868  GACCGTGAAGTCAACGACGCGCAAGTTCCTCATCAACAAACGTTCTACTTCAAGG 927
Qy      840  CTTTGAAGACGACGAGATTCCTCCGTTCTGGGGAGGAGGACCTTTTATCCATTGAT 899
Db      928  CTTTGGCAAACTAGAGACATCTCTATCAACGCGCGTGGCTTTAAGAAAGACGATGT 987
Qy      900  AAAAGACTTCAACCTTCTGAAGTGAATCAACGGAATCTTTGAGAGCTCTCAATTC 959
Db      988  GATGATTTCAATATCTCAATGATGATCGCGCAACAGCTTCGCGACGACATATTC 1047
Qy      960  TTACATGAAGTGGCTGAGATCTTGGCGACAGACTCGGAATCTTGTGATGAAGAAC 1019
Db      1048  GTACTCTAAGAGTGAATGCGTCTTGGCGATCGAGGCTGTGATCGATCGAAGAC 1107
Qy      1020  CCGGACGTTGGTAT----- 1034
Db      1108  TCCGGCAATTTGGCGTGCACCTCACTTAGCGCCACAGGAGCTCGCGAAGGACGCA 1167
Qy      1035  ---CACAGGTACCACTACATCCGAGACTCAGAAATAGCAAGAAACAACATPAAG 1091
Db      1168  GCGCGTCAAGTACTGGAGAAAGATTGGACGTTTAGACCATCAAGCGTTCTCCGTGA 1227
Qy      1092  AATGATCAAGACCAAGAACCATCCAGTGTGATCATGTGAGATGTGGGAAAGAAC 1151
Db      1228  ACTGCTGTCTGTGCAAGAACCATCAAGCGTGTGATGTGAGACCATCGCCMAAGAGC 1287
Qy      1152  AGAGTCAACCACTCAGACGCGAGGAGTTTCTTCAAGCCCTTTATGAGACTGCCATGA 1211
Db      1288  GCGCATCTAGAGAAAGGCGCGCTACGAGTACTTCAAGCGCTGTGGAGCTGACCAAG 1347
Qy      1212  AATGATGCA---ACAGCCCGCTGTGTAGTGTGATGATGACGCAACGAGAG 1268
Db      1348  ACTGCACTCAGCAAGAGCTCCGCTACGATCGTGTGTTGTGAGGCTAACCCCGAGAG 1407
Qy      1269  AACAGAGACGTGGCGCTGAAGTACTTGAATCTGTGTGAAACAGTACTAGGCTGT 1328
Db      1408  GGAACAACTCGCGCAACTGA-----TTGACCTCACTCGCGCTCAATCTCTTAAAGAG 1461
Qy      1329  GTACATCTATCAGGAGAGATGAGAGAG---GACTTCAAGCTCGAAAAAGACATAGA 1385

```

```

Db      1462  GTACTTCATGAGCGGTGATCTCGAAGCGGCCAAAGTCCATCTCCGCGAGAAATTCACGC 1521
Qy      1386  AGAGCTCTATGCAAGGCAACAGAAAGCCCATCTTTGTCAACAATTCGTTCCGAGCCAT 1445
Db      1522  GTGGAACAAGGCTGGTCCAGGAAAGCCCATATGATATCTGATGTCGCGCGCAGACCGT 1581
Qy      1446  AGCTGCATTCACATCAATCCACCTCAATGTCTCCGAAAGTGTCCAAAGGAGACTGCT 1505
Db      1582  TCGGGGCTTTCAAGCATATGATTCAGTATGTTCAAGAGAAATATCAATGTGATCTTA 1641
Qy      1506  TGAAGAAGCATCAGGCTCTTTTGAAGAAAGACTACATCAATCGAACAACAGTGTGGC 1565
Db      1642  CCAGCGCAACACAGTGTGTTGATGATGATGAGAACTTGTGGGTGAGCAAGCGTGA 1701
Qy      1566  CTTTGCAGATTTTAAAGTCTCTCAGATGTGAAAGAACCCATTCTCAACCAAGGCTGT 1625
Db      1702  CTTCCGCACTTCGAGACTCTCAGGGGTGTGATGCGCTCCAAAGGAAACAAGAGGCGT 1761
Qy      1626  TTTCAAGAGACAGACCAACCAACCTGTTGCTGATGATGAGAGACTGTGGA 1681
Db      1762  GTTCACTGTGACCGCAAGCGAGACTCGCGCGACGCTTTTCCGAGCGCTGA 1817

RESULT 6
US-09-270-957-27
; Sequence 27, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106, 405C1
; CURRENT APPLICATION NUMBER: US/09/270, 957
; NUMBER OF SEQ ID NOS: 112
; CURRENT FILING DATE: 1999-03-17
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-270-957-27

Query Match      7.7%; Score 130.4; DB 4; Length 1888;
Best Local Similarity 45.4%; Pred. No. 3.1e-30;
Matches 734; Conservative 0; Mismatches 822; Indels 61; Gaps 5;

Qy      120  CCAGGATCTGTGATGAGAAAGACCTTCACTCAAAACCACTTCTAGTCCGAA 179
Db      208  CAAGGAATTCGCAACCATATCGATATGTCTGTGACGAAGTGAATCAGGATCCGCG 267
Qy      180  GNAACCTTCAAAAAACATCAGCTTACTTGTGCGGTGAACAGGACTGCGAGT 239
Db      268  CTATCTGAAGATCAGCTATCTGTCGCTCGGCTTCCGCTCACTCAAAAGCAATTGT 327
Qy      240  CTTCTCAACGAGAGAAAGTGGAGAAATCACTTAATCTTCCCTTCTGAAATGA 299
Db      328  CTATGTCATGTGAGCTGTGTGTGAGACAAAGGCGAATCTTCCGATTTGAAAGCG 387
Qy      300  TGTGACGGGAGAAAGTGAATCCGAGAGAAAGAACTCAGGTTGTTGTGAACAGATT 359
Db      388  AATCAACAATCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 447
Qy      360  GAAAGTGGAGAGATTTCCCTCGAAGTTTCAAGACGCGCACTCAACCGTGGATTTT 419
Db      448  CGAGATAGCAACCTCCGCTGGGCTGTACAGGAGCGCACGAAGAGGCGCTTGGAAA 507
Qy      420  TGGAAATTTTCCACTGCAACTTGAATCTTCCCTAAGCGTGAATCATAGGCGCTGT 479
Db      508  AGTCAATGTGAACAAGCCGAACTTGACTTCTTCAATATGAGGCTTGCAACCGTCCGT 567
Qy      480  TCTGATAGATTCAAGACCAAGGATCTCAATCTGGGTGAGACAGAGTGAAGTC 539
Db      568  GAAATCTACAGCAGCCCGTTTACGTACGTGAGGACATCTCGTTGTGACCGACTTGA 627

```

QY 540 TGAACGGAGAGAACTTGGAAAAAGTGAAGATAGAGTCTCAGAAAGAGCGGT 599
 DB 628 TGGCCCAACCGGAGCTTGACCTTATCGGTGACTTTTCAAGCAAGCCGAGCCGTGAA 687
 QY 600 GGGACAGAGATGACGATCAAACTTGGAGAGAAAGAAAAATTAGACATCCAAAG 659
 DB 688 AGTGCCTGCTGATGAGAAAGGCAAGGTGTGCAAGCAACGAGGGCTGAGCGGTAA 747
 QY 660 ATTGCTGCAAGGGAGTTCATCTTC-GAAAAAGCCAGTTTGGAGCTCGAAGATCCAT 718
 DB 748 CGTGAAGATTCGGAATGTCATCTCTGGGAAACACTGAACAGTATCTTACCCAGATCA 807
 QY 719 ATCTTATCTCTCAAGGTGGAACTTGAAGAAAGAGTACACTCTGGACCTGGAAATCA 778
 DB 808 AAGTGAACCTGTGAAACGACGAGCTGACCATGATGTCTTAGAAGAGCCGTTCGGCGTGC 867
 QY 779 GAACGATCAGCTGGAGCAGAAAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAG 838
 DB 868 GGACCGTGAAGTCAACGACGGCAAGTCTCATCAACAAACCGTTCTACTTCAAGG 927
 QY 839 GCTTTGAAGCAGAGGAATTCCTCCGTCTGGGGCAGGGGACCTTTTATCATTTGATGA 898
 DB 928 GCTTGGCAACATGAGGACACTCTTATCAACGCGCGTGGCTTTAACGAAGCGCAATG 987
 QY 899 TAAAGACTTCAACCTTCTGAAGTGAATCAACGCGAAATCTTTGAGAGCTCTCATATC 958
 DB 988 TGATGATTTCAATATCTCTAAATGAGATCGGCGCAACAGTTCCGACCGCACATATC 1047
 QY 959 CTTCAGTGAAGATGGCTGATCTTGGCGACAGACTCGGAATCTTGTGATAGCAAG 1018
 DB 1048 CGTACTCGAAGATTTGATGCTGTGGGATTCGCGAGGGTCTGTGATCAACGAGA 1107
 QY 1019 CCGCGACGTTGGTAT----- 1034
 DB 1108 CTCCGGGAGTTGGGTGACCTCAACTTCATGCGCACACGCGAGCTGGCGAAGCGACG 1167
 QY 1035 ---CACAGGTACCACTCAATCCGAGACTCAGAAAGATAGCAAGAACATTAAGAA 1090
 DB 1168 AGCCGCTCAGTACCTGGGAGAAAGTGGAGCTTTGAGCACCATCAAGAGCTTCCGTG 1227
 QY 1091 GAATGATCGACAGACACAGAAACCATCCCACTGATGATCATGTGAGTGTGCCAAGCAAC 1150
 DB 1228 AACTGGTGTCTGTGACAGAAACCATCCAAAGCTGTGATGTGAGATGCGCCAAACAGG 1287
 QY 1151 CAGAGTCCAAACCATCGACAGCGGAGGTTCTTCAAGCCCTTATGAGATGCCAATG 1210
 DB 1288 CGGCGACTGAGGAGAGAGGCGCGCTACAGTACTTCAAGCCGTGTGAGCTGACCAAG 1347
 QY 1211 AATGATCGA---ACAAGCCCGTGTGTATGTGAGCATGATGAGCGACCAAGACAGA 1267
 DB 1348 AACTCGACCCACAGAACGTCGCGTCAAGTGTGATGATGGTATACCCCGGAGA 1407
 QY 1268 GAACAAAGAGCGTGGCGGTAGTACTTCAGCATCGTGTGTGAGAAAGATCTACGCGCT 1327
 DB 1408 CGGACAAAGTGGCCAACTGA-----TTAGCGTCAATCGCTCAATACGCTATTAACGAT 1461
 QY 1328 GGTATCATCTATCAGGGAAGATAGAAAG---GACTTCAAGCTCTTGAAGAAAGACATAG 1384
 DB 1462 GGTACTTCTGATGGGATGATCTGAAAGCGGCAAAAGTCAATCTCCGCAAGAAATTTACAG 1521
 QY 1385 AAGAGCTCTATGCAAGGCAAGAAAGCCATCTTTGTCAAGAAATTCGTGCGAGCGCA 1444
 DB 1522 CGTGAACAAAGCGTTTCCAGGAAGCCGATCATGATCACTGATAGCGGCGCAGACACCG 1581
 QY 1445 TAGCTGACATCACTATGATCAACCTCAAAATGTTCTCGAAGAGTACCAAGCAGAGCTCG 1504
 DB 1582 TTGCGGGCTTTCAGCAATGATTCAGTATGTTCCAGGAAATATCAAGTGAAGTACT 1641
 QY 1505 TTGAAAAGAGATCAGGCTCTTTTGAAGAAAGATCAATCATCTGAAACACAGTGTGG 1564
 DB 1642 ACCAGGGAGAACAGTGTGTGATGATGTTTGAAGAACTTGTGTGTGAGCAAGCGGTGA 1701

QY 1565 CTTTGCAGATTTTAAGCTCTCAGATGTGAGAAAGCACCATTCTCAACCAAGGCTG 1624
 DB 1702 ACTTCGGGAGCTTCGGGACCTCTCAGAGGCGGTGAGCGCTCAAGAAACAAAGGCGC 1761
 QY 1625 TTTTCAAGAGACAGACAAACCAACCTGTTGCTCATGTATGAGAAAGCTGTGA 1681
 DB 1762 TGTTCACGTGTGACCGCAAGCCGAGACCTGCGCGGACACGTCTTCCGAGCGCTGA 1818
 RESULT 7
 US-08-819-866-1/c
 ; Sequence 1, Application US/08819866
 ; Patent No. 5830698
 GENERAL INFORMATION:
 APPLICANT: REEF, Mitchell E.
 APPLICANT: BARRETT, Richard Spence
 APPLICANT: MCLACHLAN, Karen Retla
 TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT
 SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS
 RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSER: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
 STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/819,866
 FILING DATE: 14-MAR-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Teekin, Robin L.
 REGISTRATION NUMBER: 35,030
 REFERENCE/DOCKET NUMBER: 012712-352
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620
 TELEFAX: (703) 836-2021
 INFORMATION FOR SEQ. ID NO. 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 14683 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-819-866-1
 Query Match 6.6%; Score 112; DB 2; Length 14683;
 Best Local Similarity 48.1%; Pred. No. 6.7e-24;
 Matches 472; Conservative 0; Mismatches 450; Indels 60; Gaps 3;

QY 757 TACACTCGGACATGGAATGAAAGATCACTGGGACGAGAGAGCTCATATCGAAC 816
 DB 7348 TACCGCTTCGGGTGGCATCCGTCAGTGGCAGTGAAGGCGCAACAGTTCTGATTAAC 7289
 QY 817 GGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGGAATCCCGTTCTGGGGCAG 876
 DB 7288 CACAAACGCTTACTTATCTGCTTTGGTGTGATGAAGATGGGAACTTGTGCGGCAA 7229
 QY 877 GGCACTTTATTCATGTGATTAAGACTTGAACCTTTGAGATGATCAACCGAAT 936
 DB 7228 GGATTCGAATGAAGTGTGATGACGACCAACGATTAAATGATGATGATTTAGGCGCAAC 7169
 QY 937 TCTTTCAGGACCTTCATCTTACAGTGAAGATGGCTGATCTTGGCGAGACATC 996
 DB 7168 TCTTACGTACTTCGATTAACCTTACCTGTAAGAGATGCTGACTGGGAGATGAACAT 7109

QY 997 GGAATCCTTGATAGAGAGCC-----GCAGTGGT 1032
DB 7108 GGCACTGATGATGATGAACTGCTGCTGCTTAACTCTTTAGGCAATGGT 7049
QY 1033 ATCAACAGGTAACCACTAATCCGAGACTCAGAAATAGCAGAAAC----- 1083
DB 7048 TTCGAAGGGGGAACAGCCGGAAGAACTGTACACGAAGAGAGTCAACGGGGAAC 6989
QY 1084 -----ATAAGAAATGATGACAGACACAGAACCATCCAGT 1122
DB 6988 CACCAAGCGCACTTACAGCGCATTTAAAGCTGATAGCCGTCGACAAAACCAACCAAGC 6929
QY 1123 GGGATCATGTGGAGTGGCGGAACGAACGAGTCCACATCCAGCGGAGGTTTC 1182
DB 6928 GTTGATGATGAGATGATGACCAACGATACCCGTCGCAAGGTGCAAGGAAAT 6869
QY 1183 TTCAAAGCCCTTTATGAGACTGCAATGAAATGATCCGACAGCCGCTGTCATG 1242
DB 6868 TTCCGCGCACTGGCGGAAGCAACGGGTAACTCGACCGGCTCCATCCTGCGTC 6809
QY 1243 AGCATGATGACGACACGACGAGAAACAGAGAGTGGCGCTGAAGTCTTCACATC 1302
DB 6808 AATGTAATGTTCTGCGAGCTCACACCGATACATCAGCATCT-----CTTGATG 6755
QY 1303 GTCGTGTGAAACAGTACTAGCGCTGTATCATCTATCAGGAGATGAAAGACTT 1362
DB 6754 CTGTGCTGAACCGTTATTAACGATGATGATGTCAAAGCGGATTTGAAACGCAAG 6695
QY 1363 CAAGCTCTGAAAAAGATAGAAAGCTCTATGCAAGGACAGAAAGCCCTTTGTC 1422
DB 6694 AAGGTAAGTGAAGAAAGAACTTGTGCTGCGGAGGAAATGATCAGCGCATTTATC 6635
QY 1423 ACAGATTCGTGCGGAGCGGATAGCTGATCTCACTAGATTCACCTTAAATGTTCC 1482
DB 6634 ACCGATACCGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 6575
QY 1483 GAAGATCAACAGAGAGCTGTGAAAGAGATCAGGCTCTTTGAAAAAAGATAC 1542
DB 6574 GAAAGATACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 6515
QY 1543 ATCATCGGAACACAGTGTGCGCTTTGAGATTTTAAAGACTCTCAGAAATGAGAA 1602
DB 6514 GTGCTGCTGAACAGAGTATGAAATTTGCGCATTTTGGGACCTCCGAGGCAATG 6455
QY 1603 CCCATTTCTCAACACAGAGGTTTTCACAGAGACAGAACCAAACTGTTCTAT 1662
DB 6454 GTTGCCTGTAACAGAAAGGATCTTCACTCGGACCCGAAACGAGTGGCGGCTTT 6395
QY 1663 GTACTGAGAGCTGTGAGTG 1684
DB 6394 CTGCTGCAAAAACGCTGAGCTG 6373

RESULT 8
US-09-023-715-1/c
Sequence 1, Application US/09023715
Patent No. 598144
GENERAL INFORMATION:
APPLICANT: REFF, Mitchell E.
APPLICANT: BARNETT, Richard Spence
APPLICANT: MCLACHLAN, Karen Retta
TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT
TITLE OF INVENTION: SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS
RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,715
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/819,866
FILING DATE: 14-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 14683 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-023-715-1

Query Match 6.6%; Score 112; DB 2; Length 14683;
Best Local Similarity 48.1%; Pred. No. 6,7e-24;
Matches 472; Conservative 0; Mismatches 450; Indels 60; Gaps 3;

QY 757 TACACTTGGACATGGAATCAGAAAGATCAGCTGGACGAGAGAGAGCTTATCTGAA 816
DB 7348 TACCGCTTCGCGTCGATCGGTCAGTGGCAGTGAAGGCGAACAGTTCCTGATTA 7289
QY 817 GGGAACTGTCTTTTGAAGGCTTTGAAAGACGAGAAATCCCGTCTGAGGAG 876
DB 7288 CACAAACGTTCTTACTTACTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 7229
QY 877 GGCACCTTTATCCATTGATGATTAAGACTTCAACCTTCTGAGTGCATCAACCGAT 936
DB 7228 GGATTCGATTAACGCTGATGATGATGATGATGATGATGATGATGATGATGATG 7169
QY 937 TCTTTCAGAGCTCTCATATCTTTACAGTGAAGTGGCTGATCTTGGCAGACTC 996
DB 7168 TCTTACCTTACTCTCATATCTTACCTTACCTGAGAGATGCTGACTGGGAGATGA 7109
QY 997 GGAATCCTTGATGAGAGCC-----GCAGTGGT 1032
DB 7108 GGCACTGATGATGATGAACTGCTGCTGCTTAACTCTTTAGGCAATGGT 7049
QY 1033 ATCAACAGGTAACCACTAATCCGAGACTCAGAAATAGCAGAAAC----- 1083
DB 7048 TTCGAAGGGGGAACAGCCGGAAGAACTGTACACGAAGAGAGTCAACGGGGAAC 6989
QY 1084 -----ATAAGAAATGATGACAGACACAGAACCATCCAGT 1122
DB 6988 CACCAAGCGCACTTACAGCGCATTTAAAGCTGATAGCCGTCGACAAAACCAACCAAGC 6929
QY 1123 GGGATCATGTGGAGTGGCGGAACGAACGAGTCCACATCCAGCGGAGGTTTC 1182
DB 6928 GTTGATGATGAGATGATGACCAACGATACCCGTCGCAAGGTGCAAGGAAAT 6869
QY 1183 TTCAAAGCCCTTTATGAGACTGCAATGAAATGATCCGACAGCCGCTGTCATG 1242
DB 6868 TTCCGCGCACTGGCGGAAGCAACGGGTAACTCGACCGGCTCCATCCTGCGTC 6809
QY 1243 AGCATGATGACGACACGACGAGAAACAGAGAGTGGCGCTGAAGTCTTCACATC 1302
DB 6808 AATGTAATGTTCTGCGAGCTCACACCGATACATCAGCATCT-----CTTGATG 6755
QY 1303 GTCGTGTGAAACAGTACTAGCGCTGTATCATCTATCAGGAGATGAAAGACTT 1362

Db	6754	CTGTCCTGAAACCGTTATTACGATGTATGTCAAAGCGCGATTGGAAAACGGCAAG	6699
QY	1363	CAAGCTCTGGAAAAAGACATAGAAAGACTCTATGCAAGGCACAGAAAGCCATCTTTGTC	1422
Db	6694	AAGGTACTGGAAAAAGAACTTCTTGCCCTGGCAGGAGAACTGCACTACGCCGATTATCATTC	6639
QY	1423	ACAGATTTCGGTGGCGAGCGCATAGCTGGCATCCATCGATCCACTCCAAATGTTCTCC	1483
Db	6634	ACCGAATACGGCGGTGGATACGTTTACCGGGCTGCACCTCAATGTACACCGACATGGGAGT	6575
QY	1483	GAAGAGTACCAAGCAGCAGACTCGTTGAAAAAGACGATCAGGCTCTTTTGAAAAAAGACTAC	1542
Db	6574	GAAAGGTATCAGTGTGCATGGGTGGATATGATATACCGCGCTTTTGATCGGGTCAGCGCC	6515
QY	1543	ATCATCGGAACAACGCTGTGGCCCTTTGCAAGATTTTAAAGCTCTCAGAAATGTGAGAGA	1602
Db	6514	GTCGTCGGGTGAACAGGTATGGAAATTTCCCGATTTTGGCACTGGCAAGGCAATTTGGCC	6455
QY	1603	CCCATTTCTCAACCAAGAGGTGTTTTCACAAAGACAGACAAACCCAACTGTTGCTCAT	1662
Db	6454	GTTGGCCGGTAAACAAGAAAGGAGATCTTCACTCGCACCGCAAAACGAAAGTCCGCGGCTTTT	6399
QY	1663	GTACTGAGAAAGACTGTGGAGTG	1684
Db	6394	CTGCTGCAAAAACCTGGACTG	6373

RESULT 9

```

: Sequence 1 Application US/09343485A
: Patent No. 6413777
:
: GENERAL INFORMATION:
:
: APPLICANT: REFF, MITCHELL R.
: APPLICANT: BARNETT, RICHARD S.
: APPLICANT: MCILACHLAN, KAREN S.
: TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN
: TITLE OF INVENTION: MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND
: TITLE OF INVENTION: VECTORS FOR ACCOMPLISHING THE SAME
: FILE REFERENCE: 037003-0275807
:
: CURRENT APPLICATION NUMBER: US/09/343,485A
: CURRENT FILING DATE: 1999-06-30
:
: PRIOR APPLICATION NUMBER: 09/023,715
: PRIOR FILING DATE: 1998-02-13
:
: PRIOR APPLICATION NUMBER: 08/819,866
: PRIOR FILING DATE: 1997-03-14
:
: NUMBER OF SEQ ID NOS: 3
:
: SOFTWARE: Patentin Ver. 2.1
:
: SEQ ID NO 1
:
: LENGTH: 14683
:
: TYPE: DNA
:
: ORGANISM: Artificial Sequence
:
: FEATURE:
:
: OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
:
: IS-09-343-485A-1

```

Query Match	6.6%;	Score 112;	DB 4;	Length 14683;
Best Local Similarity	48.1%;	Pred. No. 6.7e-24;		
Matches 472;	Conservative	0;	Mismatches 450;	Indels 60;
				Gaps 3;

QY	757	TACACTCTGGACATCGGAATCAGAACGATCAGCTCTGGAACGAGAAAGGCTCTATCTTGAAC	816
Db	7348	TACCCGCTCCGCTGGGATCCGGATCGCATGGACATGAAAGGGCGCAACATTCCTGTATTAAC	7288
QY	817	GGGAAACCTGTCCTTTTGAAGGGCTTTGGAAAGCAGAGGAATTCGCCGTCCTGGGGCAG	876
Db	7288	CACAAACCGTCTCTACTTACTGGCTTTTGTCGTATGAAGATCGGACTTTGGCTGGCAAA	7229
QY	877	GGCACCTTTATCCACTTGTATGATTAAGACCTTCAACCTTCTGAAATGATCAACGCGAAT	936
Db	7228	GGATTGATTAACGTGCTGATGGTGCACAGCACCGATTAATGACTGGATTAGGGCCAAAC	7169
QY	937	TCTTTCAAGACCTCTCACTATCTTTACAGTGAAGAGTGGCTGGATCTTGGCCGACAGACTC	996

Db	7168	TCCTACGGTACCTGGCATTTACCTTTACGCTGAAGAGTCTCGACTGGGCAAGATGAACAT	7105
OY	997	GGAAATCCTTGTGATGACGAAGCCCC-----GCAAGTTGGT	1032
Db	7108	GGCATTCCGTGGTGAATGATGAAGAACTGCTGCTGCTGCGCTTTAACTCTCTTTAGGCAATTGGT	7045
OY	1033	ATCCAAAGSTACCACTCAATCCCGAGACTCAGAAAGATAGCAAGAAAGAAC-----	1085
Db	7048	TTCCGAAGCGGCGCAACAGCCGAAAGAACCTGTACACGAAGAGGAGATCCAAAGGGGAAACT	6988
OY	1084	-----ATAGAAGATGATATCGACACACACAAGAACCATCTCCAGT	1122
Db	6988	CACCAAGCGCACTTACAGGCGATTTAAAGAGCTGTATAGCGCTGACAAAACCAACCCAAAGC	6928
OY	1123	GTGATCATGTGAGGTGTGGCGCAAGAAACAGAGTCAAACATCCAGACGCGGAGGGTTTC	1182
Db	6928	GTGGTGAATGTGAAGTATTTGCCAACGAACCGGATACCCGTCGCAAGGTGACAGGGAAATAT	6865
OY	1183	TTCAAAGCCCTTTATGAGACTGCCAATGAATGATGAAACAGCCCCGTTGTATGGTG	1242
Db	6868	TTCCGCCCACTGGGCGGAAGCAACCGTAAACTCCACCCGACGCTCGATCACTCGGCTC	6809
OY	1243	AGCATGATGACGCGACCAAGACGAAGAAACAAGAACGTGGGCGTGAAGTACTTGACATC	1302
Db	6808	AATGTAATGTTCTTGGCAGCGCTCACACCGATACATCAGCGATCT-----CTTTGATGTG	6755
OY	1303	GTCGTGTGTAACAGGTACTACGCGCTGTATCATCTATCAGGGAAGAGATAGAAGAGACTT	1362
Db	6754	CTGTGCGTGAACCGTTATTACGGAATGTATGTCCAAAGCCGCAATTTGGAAACGCGACAG	6655
OY	1363	CAAGCTCTGGAAAAAGACATAGAAAGACTTATGCAAGGACAGAAAGCCCATCTTTGTC	1422
Db	6694	AAGGTACTGGAAAAAGAACTTCTGGCTCGCAGGAGAAATGCAATCAGCGGATATCATC	6635
OY	1423	ACAGAATTCGGTGGGACGGATAGCTGGGATCCACTACGATCCACTCAATATGTTTCC	1482
Db	6634	ACCGAAATACGGCGGTGAATGATTAGCCGGGCTGCATCAATGTACACCGACATGTGAGT	6575
OY	1483	GAAAGATACCAAGCAGAGCTCGTTGAAATAACGATCAGGCTCCTTTTGAAAAAAGACTAC	1542
Db	6574	GAAAGATATCAGTGTGATGGCTGGAATATGTATATCACCGGCTCTTTGATCGCGTACGGCC	6515
OY	1543	ATCATCGGAACAACGCTGTGGGCTTTTGCAATTTTAAAGCTCCTCAGAAATGTAGAGA	1602
Db	6514	GTGCTCGGTGAACAGATATGAAATTTGCGCGATTTTGGCAGCTCGAAAGCATATTTGGC	6455
OY	1603	CCCATTTCTCAACCAAGAGGTGTTTTCACAAAGACAGACACCAATCTCGTTGCTCAT	1662
Db	6454	GTTGGCGGTACAAAGAAAGGAATCTTCACTCGGACCGCAAAACCGAAGTGGCGGCTTTT	6395
OY	1663	GTACTGAGAAGACTGTGAGATG 1684	
Db	6394	CTGCTGCAAAAACGCTGGAATG 6373	

RESULT 10

US-08-723-624-18
Sequence 18, Application US/08723624
Patent No. 5861277
GENERAL INFORMATION:
APPLICANT: Rose, Alan B.
APPLICANT: Lael, Robert L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: ARNOLD, WHITE & DURKEE
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,624
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: BRIP:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
FAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-723-624-18

Query Match 6.6%; Score 110.8; DB 2; Length 3035;
Best Local Similarity 47.4%; Pred. No. 5.9e-24;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;
QY 757 TACACTCTGCATCGAATCAGATCAGCTGGAGCAGAGAGAGGCTTATCTGAAC 816
DB 2016 TACCCGCTTCGGGTGGCATCCGTCAGTGGCAGTGAAGGGGAAACAGTCTGATTAAC 2075
QY 817 GGGAACTGTCTTTTGAAGGGCTTTGAAAGCAGAGAAATCCCGTTGCGGCGAG 876
DB 2076 CACAAACCGTCTACTTACTGCTTGTGCTCATGAAGTCCGACCTTGGCGGCAAA 2135
QY 877 GGACACTTTTATCCATTGATGATTAAGACTTCAACCTTGAATGATCAACCGGAT 936
DB 2136 GGATTCGATTAACGTCTGATGATGACACGACGATTAATGACTGATTTGGGCGCAAC 2195
QY 937 TCTTCAGAGACTCTCACTATCTTAAAGAGAGAGTGGCTGATCTTGGCGAGACTC 996
DB 2196 TCCTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTGACTGGGCGAGATGAAC 2255
QY 997 GGAATCTTGTGATAGCAGAGCCCGGACGTTGTATCAAAAGT----- 1042
DB 2256 GGCAATCGTGTGATGATGAAGCTGTGCTGCGCTTAACTCTCTTAGGCAATGGT 2315
QY 1043 -----ACCACTACAAATCCGAGACT 1062
DB 2316 TTCGAAGCGGCGAACAGCCGAAAGACTGTACAGCGAAGAGGAGTCAACGGGAAACT 2375
QY 1063 CAGAAATAGCAGAGACAAACATTAAGAAATGATGACAGACAGACAAAGCATCCAGT 1122
DB 2376 CAGCAAGGCACTTACAGCCGATTAAAGAGCTGATAGCCGCGTGAACAAACCAACCAAGC 2435
QY 1123 GTGATCATGTGAGTGTGGCGAAGCAGAGTCAACCATCTCAAGACCGGAGGGTTTC 1182
DB 2436 GTGGATGTGGAGATGATGCCAACAGACCGGATACCCGTCGCAAGGTGCAAGGAAATAT 2495
QY 1183 TTCAAAGCCCTTATAGAGCTGCCAATGAATGATCGAACCGCCCGTGTCTATGTG 1242
DB 2496 TTCGCGCACTGGCGGAAAGCAACGGCTAACTCGACCGACGCTCGATCACTGCGTTC 2555
QY 1243 AGCATGATGACGACACAGACAGAGAAACAGAGAGTGGCCGCTGAAGTCTTGAACATC 1302
DB 2556 AATGTAATGTTTGGAGAGCTCAACACCATCAAGAGATCT-----CTTTATGTG 2609
QY 1303 GTCTGTGTAACAGGTAACGCTGCTGATCATCTATCAGGGAAGATGAGAAAGAGACTT 1362
DB 2610 CTGTGCTGTAACCGTTATTAACGATGATATGTCAAAGGGGAGATTTGAAACGCGAAG 2669
QY 1363 CAAGCTCTGGAAGAAAGATAGAGAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTC 1422

DB 2670 AAGTACTGGAAGAAAGACTTCTGCGTCGAGAGAAATGATCAAGCCGATTTATCATC 2729
QY 1423 ACAGAAATTCGCTGCGAGACCGATAGCTGCAATCCATCAATCAATGTTCTCC 1482
DB 2730 ACCGAATACGGGCTGATACGTTAGCCGGGCTGCACTCAATGTACACCGACATGTGAGT 2789
QY 1483 GAAAGTACCAAGCAGAGCTGTTAAAGACGATCAGGCTCTTTGAAAAAGACTAC 1542
DB 2790 GAAAGTATCAGTGTGATGCTGTGATATGATCAACCGGCTTTGATCGCGTCAGCCG 2849
QY 1543 ATCATCGGAACACAGCTGGGCTTTGCAATTTTAAGACTCTCAGAAATGTGAGAGA 1602
DB 2850 GTCTGCGTGAACAGATATGAAATTTCCCGATTTTGGGACCTCCGAAAGCATATTCGC 2909
QY 1603 CCATCTTCACCAACAGGCTGTTTCAAGAGACAGACCAACCACTCGTGTCTAT 1662
DB 2910 GTTGGCGGTAACAAAGAGGATCTTCACTGCGGACCGCAACCGAAGTGGCGGCTTTT 2969
QY 1663 GTACTGAGAAAGACTGTGAGTG 1684
DB 2970 CTGCTGCAAAAACGCTGACTG 2991

RESULT 11
US-08-630-820-5
Sequence 5, Application US/08630820
Patent No. 6008023
GENERAL INFORMATION:
APPLICANT: OPPER, Martin
APPLICANT: BOSSLET, Klaus
APPLICANT: CZECH, Joerg
TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,820
FILING DATE: 10-APR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 19513676.4
FILING DATE: 11-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: GRANADOS, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 18748/306
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3169 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORGANISM: Enterobacteriaceae: Escherichia coli

Best Local Similarity 47.4%; Pred. No. 6e-24;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```
QY 757 TACACTGACATGGAATGAAAGATGATGAGTGGAGGAGAGAGAGGCTTATCTGAAC 816
Db 2145 TACCCGCTTCGGGTGGGATCCGGTCAAGGCGAGTGAAGGGGGAACAGTTCTGATTAAAC 2204
QY 817 GGGAAACCTGCTTTTGAAGGCTTTGAAAGCAAGAGATTCCTCCGTTCTGGGGCAG 876
Db 2205 CACAAACCGTTCTTACTTACTGCTTGGCTTGGCTCTCAATGAAGAGCGGACTTACGTGCAAA 2264
QY 877 GGCACCTTTTATCCATGATGATTAAGACTTCAACTTTGATGATGATCAACCGGAT 936
Db 2265 GGATTGATTAACGTGCTATGCTGACACACGATTAATGATGATGATGGGCGCAAC 2324
QY 937 TCTTTGAGACCTCTCATCTTCTTACAGTGAAGAGTGGCTGATCTTGGCCGACACTC 996
Db 2325 TCTTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTGACTGGGCAAGATGAACAT 2384
QY 997 GGAATCCTTGATGATGAGGAAGCCCCGACGTTGGTATCAAAAGT----- 1042
Db 2385 GGCATCGTGTGATGATGAAGAACTGCTGCTGCGCTTTAACCTCTCTTTAGGCAATTGGT 2444
QY 1043 -----ACCATCAATCCGAGACT 1062
Db 2445 TTCCGAAGCGGGCAACAAAGCCGAAAGAACTGTACAGCAAGAGGCGACTCAAGGGGAAACT 2504
QY 1063 CAGAAATAGACAGAAAGACAACTAAAGAAATGATGACAGACAGACAGAACCATCCAGT 1122
Db 2505 CAGCAAGCGCACTTACAGGCGATTAAGAGCTGATAGCGCGTGAACAAAACACCCAAAGC 2564
QY 1123 GTGATCATGTGAGTGTGGCGCAACGAACTCAAGTCCAACTCCAGACGGCGAGGTTTC 1182
Db 2565 GTGGATGTGAGTGTGGCGCAACGAACTCCGCTCCGCAAGGTGACAGGAAATAT 2624
QY 1183 TTCAAAGCCCTTTATGAGACTGCCAATGAATGATGCAACAGCCGCTTGTGATGGTG 1242
Db 2625 TTCGGCCACTGGCGGAGCAAGCGGCTAACTCGAACCGGCTCCATCATCTGCGTC 2684
QY 1243 AGCATGATGACGCAACGACGAGAAACAGAGAGCTGGCGCTGAAGTACTTGCACATC 1302
Db 2685 AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGGATCT-----CTTTGATGTG 2738
QY 1303 GTCTGTGTGAACAGTACTACGCGCTGTGATCTATCAGGGAAGATAGAAAGAGACTT 1362
Db 2739 CTGTGCTGAACCGTTATTAACGATGTATGTCCAAAGCGGATTTGAAACGGCAGAG 2798
QY 1363 CAAGCTGTGAAAAAGACATAGAAAGAGCTGTATGCAAGGACAGAAAGCCATCTTTGTC 1422
Db 2799 AAGGTACTGAAAAAGAACTTGTGCTGTGGCAGGAGAAATGCTCATGACCCATTTATCATC 2858
QY 1423 ACAGAAATTCGATGCGGACGATAGCTGGCATCTACATGATCCACTCAATGTTCTCC 1482
Db 2859 ACCGAAATTCGCGGTGATACGTTAGCCGCGGTGCTCATATGTAACCGGACATGTGGAGT 2918
QY 1483 GAAAGTACCAAGCAGAGCTGTTGAAAAAGACATCAGGCTCTTTTGAAGAAAGACTAC 1542
Db 2919 GAAAGTATCAGTGTGATGCTGTGATGTATGTACACCGGCTCTTTGATGCGCTGACGCC 2978
QY 1543 ATCATCGGAACACAGTGGGCTTTGAGATTTTAAAGACTCTCAGAAATGTGAGAA 1602
Db 2979 GTCTGCGGTGAACAGATATGAAATTTGCGCAATTTTGGCACTTCGCAAGGCAATATGGGC 3038
QY 1603 CCCATTCTCAACCAAGAGGTGTTTTCAGAGAGACAGACAAACCCAACTCGTTGCTCAT 1662
Db 3039 GTTGGCGGTAAACAAAGAGGATCTTCACTGCGACCGCAACCGAAGTGGCGGCTTTT 3098
QY 1663 GTACTGAGAAAGACTGTGAGAGT 1684
Db 3099 CTGCTGCAAAAACGCTGAGCTG 3120
```

RESULT 13

US-08-723-624-19
; Sequence 19, Application US/08723624
; Patent No. 3861277

```
; GENERAL INFORMATION:  
; APPLICANT: Rose, Alan B.  
; APPLICANT: Last, Robert L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING  
; TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ARNOLD, WHITE & DURKEE  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,624  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: BTIP:002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 512/418-3000  
; TELEFAX: 512/474-7577  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3824 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-723-624-19
```

Query Match 6.6%; Score 110.8; DB 2; Length 3824;

Best Local Similarity 47.4%; Pred. No. 6.8e-24;

Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```
QY 757 TACACTGACATGGAATGAAAGATGATGAGTGGAGGAGAGAGGCTTATCTGAAC 816
Db 2805 TACCCGCTTCGGGTGGGATCCGGTCAAGTGAAGAGGGGGAACAGTTCTGATTAAAC 2864
QY 817 GGGAAACCTGCTTTTGAAGGCTTTGAAAGCAAGAGATTCCTCCGTTCTGGGGCAG 876
Db 2865 CACAAACCGTTCTTACTTACTGCTTGGCTGTGCTCAATGAAGATGCGGACTTGGTGGCAAA 2924
QY 877 GGCACCTTTTATCCATGATGATTAAGACTTCAACTTTGAAATGATGATCAACCGGAT 936
Db 2925 GGATTGATTAACGTGCTATGCTGACACACGATTAATGATGATGATGGGCGCAAC 2984
QY 937 TCTTTGAGACCTCTCATCTTCTTACAGTGAAGAGTGGCTGATCTTGGCCGACACTC 996
Db 2985 TCTTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTGACTGGGCAAGATGAACAT 3044
QY 997 GGAATCCTTGATGATGAGGAAGCCCCGACGTTGGTATCAAAAGT----- 1042
Db 3045 GGCATCGTGTGATGATGAAGAACTGCTGCTGCTTAACTCTCTTTAGGCAATTGGT 3104
QY 1043 -----ACCATCAATCCGAGACT 1062
Db 3105 TTCCGAAGCGGGCAACAAAGCCGAAAGAACTGTACAGCGGAAGAGGCACTGAACGGGGAAC 3164
QY 1063 CAGAAATAGACAGAAAGACAACTAAAGAAATGATGACAGACAGACAGAACCATCCAGT 1122
Db 3165 CAGCAAGCGCACTTACAGGCGATTAAGAGAGCTGATAGCGCGTGAACAAAACACCCAAAGC 3224
QY 1123 GTGATCATGTGAGTGTGGCGAAGAACCAAGTCCAACTCCAGACGGCGAGGTTTC 1182
```

Db 3225 GTGGTATGTGGATTTGCCAAGAACCGGATACCCGTCGCCAAGGTGACGCGAATAT 3284
 Qy 1183 TTCMAAGCCCTTTATGAGACTGCGCAATGAAATGATGAAACGCCCGGTGTGCTG 1242
 Db 3285 TTGCGGCCACTGCGGAAACCAACGGTAAATCGACCCGACCGGTCCGATCACTGCGTC 3344
 Qy 1243 AGCATGATGAGCAACAGACGAGAAACAGAACAGTGGCGCTGAATGATCTGACATC 1302
 Db 3345 AATGTAATTTCTGGAGCGCTCAACCGATACCATCAGCGATCT-----CTTTGATGTG 3398
 Qy 1303 GTCTGTGTAACAGGTAAGTACGCGCTGTGTAATCTATCAGGAAAGATAGAAAGACTT 1362
 Db 3399 CTGTGCTGTAACCGTTATTAACGATGTGTGTCAAAGGCGGATTTGAAACGCGCAG 3458
 Qy 1363 CAAGCTCTGGAAGAAAGACATAGAAAGCTCTTTCAGAAAGCAGACAAAGCCATCTTTC 1422
 Db 3459 AAGGTAATGAAAGAAAGAACTTCTGCTGCGAGGAAACCTGCATCAACCGATTAATCATC 3518
 Qy 1423 ACAGATTCGATGCGGACGCGATAGCTGATCCATCCATCCATCCATCCATCCATCCATCC 1482
 Db 3519 ACCGAATACGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3578
 Qy 1483 GAAAGTACCAAGCAAGAGCTGTTGAAAGACGATCAGCTCTTTTGAAGAAAGACTAC 1542
 Db 3579 GAAAGTATCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3638
 Qy 1543 ATCATCGGAACACAGTGTGGGCTTTGCAATTTTAACTCTCCAGAAATGTGAAGA 1602
 Db 3639 GTGCTGCGTAACAGATATGAAATTTCCGCAATTTTGGACCTCGCAAGGCAATTTGCGC 3698
 Qy 1603 CCCATTCTCAACCAAGAGGTGTTTCAAGAGACAGAACCAACCAACTGTTGCTCAT 1662
 Db 3699 GTTGGCGGTAAACAAGAAAGGATCTTCACTCGCAGCCGCAAGAGTGGCGGCTTTT 3758
 Qy 1663 GTACTGAGAAAGCTGTGAGTG 1684
 Db 3759 CTGCTGCAAAAACGCTGAGCTG 3780

RESULT 14

US-09-893-525-36

Sequence 36 Application US/09893525

Patent No. 6753167

GENERAL INFORMATION:

APPLICANT: Moloney, Maurice M.

APPLICANT: Van Rooijen, Gijb

TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies

FILE REFERENCE: 9369-172

CURRENT APPLICATION NUMBER: US/09/893,525

PRIOR FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 09/210,843

PRIOR FILING DATE: 1998-12-15

PRIOR APPLICATION NUMBER: US 08/846,021

PRIOR FILING DATE: 1997-04-25

PRIOR APPLICATION NUMBER: US 08/366,783

PRIOR FILING DATE: 1994-12-30

PRIOR APPLICATION NUMBER: US 08/142,418

PRIOR FILING DATE: 1993-11-16

PRIOR APPLICATION NUMBER: US 07/659,835

PRIOR FILING DATE: 1991-02-22

NUMBER OF SEQ ID NOS: 42

SOFTWARE: PatentIn version 3.1

SEQ ID NO 36

LENGTH: 4652

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Phae-GUS-phae

NAME/KEY: CDS

LOCATION: (1548)..(3359)

OTHER INFORMATION:

US-09-893-525-36

Query Match 6.6%; Score 110.8; DB 4; Length 4652;
 Best Local Similarity 47.4%; P-Val 7.7e-24;
 Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

Qy 757 TACCTGTGACATCGGAATCAGAACATGATGAGTGGAGCAGAAAGAGCTCTTATGAAAC 816
 Db 2343 TACCGCTGTGCGTGGCATCCGATCGAGTGCAGTGAAGGCCCAACAGTTCGATTAAC 2402
 Qy 817 GGGAAACCTGCTTTTGAAGGCTTTGAAACAGAGGAATTCCTGTTGGGCGAG 876
 Db 2403 CACAAACGTTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 2462
 Qy 877 GGCACCTTTATCATTTGATGATTAAGATTAAAGCTTAACTTCTGAAGTGAATCGCAAT 936
 Db 2463 GGAATTCATTAACGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2522
 Qy 937 TCTTTCAGAGCTCTCACTATCTCTTACAGTGAAGATGCTGTGATCTTTCGCGCAGACTC 996
 Db 2523 TCTTACCGTACCTTCGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 2582
 Qy 997 GGAATCTTGTGTAGACGAAAGCCCGCACTTGTGTATCAAGAT----- 1042
 Db 2583 GGCATCGTGTGATGATGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2642
 Qy 1043 -----ACCATTAATCCGAGCT 1062
 Db 2643 TTGAAACGGGCAACAAAGCCGAAAGAACTGTACAGGAAAGAGCAAGGCAAGGCAAGGCA 2702
 Qy 1063 CAGAAATAGCAGAAAGCAACATTAAGAAATGATGACAGACACAAAGAACATCCAGT 1122
 Db 2703 CAGCAAGCCGACTTAAGCGGATTAAGAGCTGATAGCGCGTGAACAAACCAACCAAC 2762
 Qy 1123 GTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1182
 Db 2763 GTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2822
 Qy 1183 TTCAAGCCCTTTATGAGCTGCAATGAATGAATGATGATGATGATGATGATGATGATG 1242
 Db 2823 TTGCGCCACTGCGGAGCAACCGGTAACTGACCGAGCGTCCGATCACTGCGCTC 2882
 Qy 1243 AGCATGATGAGCAACAGACGAGAAACAGAAACAGAGAGCTGAGAGTATTCAGATC 1302
 Db 2883 AATGTAATGTTCTGCGAGCTCACACCGATACCATCAGGAGATCT-----CTTTGATGTG 2936
 Qy 1303 GTCTGTGTAACAGGATCAAGCTGTGATCATCTTACAGGAAAGATGAAAGAGACTT 1362
 Db 2937 CTGTGCTGAACCGTATTAACGATGTGATGTCAAAGCGGATTTGGAACGCGCAGAG 2996
 Qy 1363 CAAGCTCTGGAAGAAAGACATAGAGAGCTCTATGCAAGCAAGAAAGCCATCTTTGTC 1422
 Db 2997 AAGTACTGGAAGAAAGAACTTTGCGCTGCGCAGAGAAATGATCAGCGCATATCATC 3056
 Qy 1423 ACAGAAATTCGTCGCGAGCGATAGCTGCAATCCATGATCACTCAATGTTCTTC 1482
 Db 3057 ACCGAATACGGGTGTGATAGTTAGCCGGGCTGCACTCAATGATACCGAGATGTGAGT 3116
 Qy 1483 GAAAGTACCAAGCAGAGCTGTTGAAGAGATCAGAGCTCTTTTGAAGAAAGACTAC 1542
 Db 3117 GAAAGTATCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3176
 Qy 1543 ATCATCGGAACACAGTGTGGGCTTTGCAATTTTAACTCTCAGATGTGAGAGA 1602
 Db 3177 GTGCTGCGTAACAGATATGAAATTTGCGCATTTTGGGCACTCGCAAGCATATGCGC 3236
 Qy 1603 CCATTTCTCAACCAAGAGGTGTTTCAAGAGACAGACAAACCAACCTGTTGCTCAT 1662
 Db 3237 GTTGGCGGTAAACAAGAAAGGATCTTCACTGCGCACCGCAACGAAAGTGGCGGCTTTT 3296
 Qy 1663 GTACTGAGAAAGCTGTGAGTG 1684
 Db 3297 CTGCTGCAAAAACGCTGAGCTG 3318

RESULT 15
 US-09-893-525-41
 ; Sequence 41, Application US/09893525
 ; Patent No. 6753167
 ; GENERAL INFORMATION:
 ; APPLICANT: Moloney, Maurice M.
 ; APPLICANT: Van Rooijen, Gtjs
 ; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
 ; FILE REFERENCE: 9369-172
 ; CURRENT APPLICATION NUMBER: US/09/893,525
 ; PRIOR FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 09/210,843
 ; PRIOR FILING DATE: 1998-12-15
 ; PRIOR APPLICATION NUMBER: US 08/846,021
 ; PRIOR FILING DATE: 1997-04-25
 ; PRIOR APPLICATION NUMBER: US 08/366,783
 ; PRIOR FILING DATE: 1994-12-30
 ; PRIOR APPLICATION NUMBER: US 08/142,418
 ; PRIOR FILING DATE: 1993-11-16
 ; PRIOR APPLICATION NUMBER: US 07/659,835
 ; NUMBER OF SEQ ID NOS: 42
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 41
 ; LENGTH: 5390
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: phas-caleo-gus-phas
 ; NAME/KEY: CDS
 ; LOCATION: (1548)..(4097)
 ; OTHER INFORMATION:
 ; US-09-893-525-41

Query Match 6.6%; Score 110.8; DB 4; Length 5390;
 Best Local Similarity 47.4%; Pred. No. 8.5e-24;
 Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY 757 TACACTTGACATGGAATGAAAGATGATGAGTGGAGGAGAGAGAGGCTTATCTGAAC 816
 DB 3081 TACCGCTTCGCGTGGGATCGGATCGAGTGGAGAGAGGCGCAAGTTCTGATTAAC 3140
 QY 817 GGGAAACCTGTTTGTGAAAGGCTTTGAAAGCAGAGAAATCCCGCTTTGGGGCAG 876
 DB 3141 CACAAACGTTCTTACTTACTGCTTGTGCTCATGAAGATGCGACTTACGTGCAAA 3200
 QY 877 GGCACCTTTTATCATGATGATTAAGACTTCAACCTTCTGAGTGAATCAACGGAAT 936
 DB 3201 GGATTCGATTAAGCTGCTGATGTCACGACGCACTTAATGACTGGATTGGGCGCAAC 3260
 QY 937 TCTTTCAGAGCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTGCCGACAGACTC 996
 DB 3261 TCCTACCGTACTCTGATTAACCTTAACGCTGAAGAGATGCTGACTGGGCAATGAACAT 3320
 QY 997 GGAATCTTGTGATGAGAGAGCCCGCAGCTTGGTATCAAGGT----- 1042
 DB 3321 GGCAATCGTGTGATGATGAAGAACTGCTGTGGCTTTTCCTCTCTTTAGGCAATGGT 3380
 QY 1043 -----ACCATCAAAATCCGAGACT 1062
 DB 3381 TTCGAAGCGGCGCAACAGCCGAAAGAACTGTACAGCGAAGAGGCACTCAACGGGAAACT 3440
 QY 1063 CAGAAGATAGCAGAAAGCACTAAAGAAATGATGACAGACACAGAAACCATCCAGT 1122
 DB 3441 CAGCAGCGCACTTACAGCGCATTAAGAGCTGATAGCGCGTGAACAAAAACACCAAGC 3500
 QY 1123 GTGATCATGTGAGTGTGCGAAAGCAAGTCCAAACATCCAGACGCGGAGGTTTC 1182
 DB 3501 GTGGATGTGAGTATGCGCAACGACCGATACCCGTCGCAAGGTGCAAGGAAATAT 3560
 QY 1183 TTCAAAGCCCTTATGAGACTGCAATGAATGATGCAACAGCCCGGTTGTCAATGGTG 1242

DB 3561 TTGCGGCCACTGGCGGAAGCAACGGTAAACTCGAACCCGAGCGCTGCCATCACTGCCGC 3620
 QY 1243 AGCATGATGAGACGACACAGACGAGAAACAGAGACGTGGCCCTGAAGTACTTTCGACATC 1302
 DB 3621 AATGTAAATGTTCTGCGAGCGCTCACACCGATACCATCAGCGATCT-----CTTGAATGTC 3674
 QY 1303 GTCTGTGGAACAGGATCACTAGCGCTGTGATCTATCTACAGGGAAGATGAAGAGGACTT 1362
 DB 3675 CTGTCCCTGAACCGTTATTTACGAGATGTATGTCGAAAGCGGAGATTGGAACGCGACAG 3734
 QY 1363 CAAGCTCTGAAAAAAGACATAGAAAGCTCTATGCAAGGCAAGAAAGCCATCTTTGTC 1422
 DB 3735 AAGTACTGGAAGAAAGAACTTGTGGCTGGCGAGGAAAGTCAATGCAATGATTTATCATC 3794
 QY 1423 ACAGATTCGGTGGCGAGCGCATAGCTGGCATCTACATGATCCACCTCAATGTTCTCC 1482
 DB 3795 ACCGAATACGGCGGTGATATGATGACCGGCTGCACCTCAATGTACACCGACATGTGAGAT 3854
 QY 1483 GAAGAGTACCAAGCAGAGCTGTTGAAAAAGATCAAGGCTCTTTGAAAAAGACTAC 1542
 DB 3855 GAAGAGTATCAAGTGTGATGCTGATATGATCAACCGGCTTTGATGCGCTCAGCGCC 3914
 QY 1543 ATCATCGGAACACAGTGTGGCTTTGCAATTTTAAGACTCTCAGATGTGAGAGA 1602
 DB 3915 GTCTGTGGTGAACAGTATGGAATTTGCGCATTTTGGCACTCCGAAAGCATATTGCGC 3974
 QY 1603 CCCATTCTCAACCAAGGAGTGTTCACAGAGACAGACCAACCAACTGTTGCTCAT 1662
 DB 3975 GTTGGCGGTAAACAAAGAGGATCTTCACTGCGACCCCAACGAAAGTGGCGGCTTTT 4034
 QY 1663 GTACTGAGAAGCTGTGAGATG 1684
 DB 4035 CTGCTCAAAAACGCTGACCTG 4056

Search completed: January 24, 2005, 07:33:15
 Job time : 170 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2005, 03:41:50 ; Search time 958 Seconds
(without alignments)
10130.277 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689

Sequence: 1 atggttaagaccgcaacga...gaagactctgagtgagctt 1689

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
6: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1688	99.9	1689	US-10-364-649-14	Sequence 14, Appl
2	1688	99.9	1692	US-10-369-493-26682	Sequence 26682, A
3	174	10.3	2169	US-10-136-841-5	Sequence 5, Appl
4	174	10.3	2169	US-10-272-531A-5	Sequence 5, Appl
5	174	10.3	2169	US-10-272-483A-5	Sequence 5, Appl
6	172.4	10.2	1956	US-10-421-175-1	Sequence 1, Appl
7	172.4	10.2	2191	US-10-388-360-343	Sequence 14, Appl
8	172.4	10.2	2191	US-10-429-802-16	Sequence 16, Appl
9	172.4	10.2	2191	US-10-430-503-7	Sequence 7, Appl
10	172.4	10.2	2191	US-10-335-053-51	Sequence 51, Appl
11	166.8	9.9	2472	US-10-388-934-225	Sequence 225, Appl
12	159.6	9.4	1947	US-10-421-175-3	Sequence 3, Appl

13	141.4	8.4	1887	US-10-120-145-7	Sequence 7, Appl
14	141.4	8.4	3651	US-10-322-656-45	Sequence 45, Appl
15	141.4	8.4	4084	US-10-322-656-48	Sequence 48, Appl
16	130.4	7.7	1888	US-10-364-649-27	Sequence 27, Appl
17	114	6.7	3451	US-10-161-403-108	Sequence 108, Appl
18	114	6.7	3451	US-10-161-408-20	Sequence 20, Appl
19	114	6.7	14627	US-10-161-403-109	Sequence 109, Appl
20	114	6.7	14627	US-10-161-408-21	Sequence 21, Appl
21	112	6.6	14683	US-10-109-853-1	Sequence 1, Appl
22	112	6.6	14683	US-10-817-950-1	Sequence 1, Appl
23	110.8	6.6	1809	US-10-432-777-16	Sequence 16, Appl
24	110.8	6.6	1812	US-10-161-403-105	Sequence 105, Appl
25	110.8	6.6	1812	US-10-369-493-24529	Sequence 24529, A
26	110.8	6.6	1812	US-10-161-408-16	Sequence 16, Appl
27	110.8	6.6	1812	US-10-149-533A-33	Sequence 33, Appl
28	110.8	6.6	2000	US-09-887-576-582	Sequence 582, Appl
29	110.8	6.6	2001	US-10-239-907A-38	Sequence 38, Appl
30	110.8	6.6	2001	US-10-332-406A-21	Sequence 21, Appl
31	110.8	6.6	2001	US-10-800-161-29	Sequence 29, Appl
32	110.8	6.6	2186	US-10-755-328-2	Sequence 2, Appl
33	110.8	6.6	2725	US-10-239-907A-43	Sequence 43, Appl
34	110.8	6.6	2730	US-10-239-907A-41	Sequence 41, Appl
35	110.8	6.6	4072	US-10-239-907A-49	Sequence 49, Appl
36	110.8	6.6	4341	US-10-239-907A-47	Sequence 47, Appl
37	110.8	6.6	4652	US-09-893-525-16	Sequence 36, Appl
38	110.8	6.6	4652	US-10-763-380-36	Sequence 36, Appl
39	110.8	6.6	5390	US-09-893-525-41	Sequence 41, Appl
40	110.8	6.6	5390	US-10-763-380-41	Sequence 41, Appl
41	110.8	6.6	5418	US-09-893-525-38	Sequence 38, Appl
42	110.8	6.6	5418	US-10-763-380-38	Sequence 38, Appl
43	110.8	6.6	5898	US-10-640-422-160	Sequence 160, Appl
44	110.8	6.6	8654	US-09-845-064-11	Sequence 11, Appl
45	110.8	6.6	8654	US-09-845-064-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-10-364-649-14
; Sequence 14, Application US/10364649
; Publication No. US20030229921A1
GENERAL INFORMATION:
APPLICANT: Richard A. Jefferson and Jorge E. Mayer
TITLE OF INVENTION: MICROBIAL B-GLUCONIDASE GENES, GENE
FILE REFERENCE: 190106.405C1
CURRENT APPLICATION NUMBER: US/10/364,649
PRIOR APPLICATION NUMBER: 2003-02-12
PRIOR FILING DATE: 2003-02-12
PRIOR APPLICATION NUMBER: US 09/270,957
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 112
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 1689
TYPE: DNA
ORGANISM: Thermotoga maritima
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1689)
OTHER INFORMATION: n = A,T,C or G
US-10-364-649-14

Query Match 99.9%; Score 1688; DB 15; Length 1689;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTTAAGACCGCAACGAAGAGATTATTTCTTATCTTGATGAGTTGGAT 60
DB 1 ATGTTAAGACCGCAACGAAGAGATTATTTCTTATCTTGATGAGTTGGAT 60

QY 61 CTGGAAGTACGAGCAAGACCAATCGCGTTCTGGAAGTGAATGAGCAATAC 120
 DB 61 CTGGAAGTACGAGCAAGACCAATCGCGTTCTGGAAGTGAATGAGCAATAC 120
 QY 121 CAGGATCTGTGTGCTACGAGAGAGAGACCTTCACTTCAAAAACCACTTCTACCTTCCGAG 180
 DB 121 CAGGATCTGTGTGCTACGAGAGAGAGACCTTCACTTCAAAAACCACTTCTACCTTCCGAG 180
 QY 181 NAACTTTCACAAAACACATCACTTACTTGTGTGTGTGAACAAGAGCTGCGAGTTC 240
 DB 181 NAACTTTCACAAAACACATCACTTACTTGTGTGTGTGAACAAGAGCTGCGAGTTC 240
 QY 241 TTCTCTCAACGAGAGAGAGAGAGAGAGATCACTTGAATACCTTCCCTCGAAGTAGAT 300
 DB 241 TTCTCTCAACGAGAGAGAGAGAGAGAGATCACTTGAATACCTTCCCTCGAAGTAGAT 300
 QY 301 GTGACGGGAG 360
 DB 301 GTGACGGGAG 360
 QY 361 AAAGTGGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGAGATTTT 420
 DB 361 AAAGTGGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGAGATTTT 420
 QY 421 GGAAGTTTTCACCTGCAAACTTTCGACTTCTTCCCTACGGTGAATCATTAAGGCTGT 480
 DB 421 GGAAGTTTTCACCTGCAAACTTTCGACTTCTTCCCTACGGTGAATCATTAAGGCTGT 480
 QY 481 CTGATAGATTCACAGACACGCGAGATTCCTGACATCTGGGTGAGACAGAGAGTCT 540
 DB 481 CTGATAGATTCACAGACACGCGAGATTCCTGACATCTGGGTGAGACAGAGAGTCT 540
 QY 541 GAACCGGAG 600
 DB 541 GAACCGGAG 600
 QY 601 GGAAGTGGAG 660
 DB 601 GGAAGTGGAG 660
 QY 661 TTCTCTCAACGAG 720
 DB 661 TTCTCTCAACGAG 720
 QY 721 CTGATAGATTCACAGACACGCGAGATTCCTGACATCTGGGTGAGACAGAGAGTCT 780
 DB 721 CTGATAGATTCACAGACACGCGAGATTCCTGACATCTGGGTGAGACAGAGAGTCT 780
 QY 781 ACGATCAGCTGAG 840
 DB 781 ACGATCAGCTGAG 840
 QY 841 TTGGAAG 900
 DB 841 TTGGAAG 900
 QY 901 AAAGACTTCAACCTTCTGAAAGTGAATCAACGCGAATCTTTCAGGACCTCTCACTATCT 960
 DB 901 AAAGACTTCAACCTTCTGAAAGTGAATCAACGCGAATCTTTCAGGACCTCTCACTATCT 960
 QY 961 TACAGTGAAG 1020
 DB 961 TACAGTGAAG 1020
 QY 1021 CCGACAGCTGTGTATCACAAGGTACCACTAATCCCGAGACTGAGAGAGAGAGAGAG 1080
 DB 1021 CCGACAGCTGTGTATCACAAGGTACCACTAATCCCGAGACTGAGAGAGAGAGAGAG 1080
 QY 1081 AACTTAAG 1140
 DB 1081 AACTTAAG 1140
 QY 1141 GCGAAG 1200

DB 1141 GCGAAG 1200
 QY 1201 ACTGCCAATGAATAGATTCGAACACGCGCTGTGTCAATGTGAGCATGATGAGACGCA 1260
 DB 1201 ACTGCCAATGAATAGATTCGAACACGCGCTGTGTCAATGTGAGCATGATGAGACGCA 1260
 QY 1261 GACGAG 1320
 DB 1261 GACGAG 1320
 QY 1321 TAGGCTGTGTATCTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 DB 1321 TAGGCTGTGTATCTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1381 ATGAAGAGCTTATGAG 1440
 DB 1381 ATGAAGAGCTTATGAG 1440
 QY 1441 GCGATAGCTGTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1500
 DB 1441 GCGATAGCTGTGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 1500
 QY 1501 CTGCTGAG 1560
 DB 1501 CTGCTGAG 1560
 QY 1561 TGGGCTGTGTGAGATTTTAAAGTCTCTGAGATGTGAGAGAGAGAGAGAGAG 1620
 DB 1561 TGGGCTGTGTGAGATTTTAAAGTCTCTGAGATGTGAGAGAGAGAGAGAGAG 1620
 QY 1621 GGTGTTTCAAG 1680
 DB 1621 GGTGTTTCAAG 1680
 QY 1681 AGTGAAGTT 1689
 DB 1681 AGTGAAGTT 1689
 RESULT 2
 US-10-369-493-26682
 ; Sequence 26682, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 26682
 ; LENGTH: 1692
 ; TYPE: DNA
 ; ORGANISM: Thermotoga maritima
 US-10-369-493-26682
 Query Match 99.9%; Score 1688; DB 15; Length 1692;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATGTAAG 60
 DB 1 ATGTAAG 60
 QY 61 CTGGAAGTACGAGCAAGACCAATCGCGTTCTGGAAGTGAATGAGCAATAC 120

Db	61	CTTGAGTGAACCGAGAAAGACAGACCAATCGCCGTTCTCTGGAAGCTGGAAATGACAGATAC	120
Qy	121	CAGAGTCTGTGCTACGAAAGAGACCTTTCACCTACAAACCACCTTCTACGTTCCGAA	180
Db	121	CAGAGTCTGTGCTACGAAAGAGACCTTTCACCTACAAACCACCTTCTACGTTCCGAA	180
Qy	181	MAACTTTTCACAAAAACATCATGACTTTTACTTGTCTGGGTGTAAACAAGCATCGCAGGCTC	240
Db	181	GAACCTTTCACAAAAACATCATGACTTTTACTTGTCTGGGTGTAAACAAGCATCGCAGGCTC	240
Qy	241	TTCTCTCAACGAGAGAAAAGTGGAGAGATCACTTGAATACCTTCCCTCGAAGTAGAT	300
Db	241	TTCTCTCAACGAGAGAAAAGTGGAGAGATCACTTGAATACCTTCCCTCGAAGTAGAT	300
Qy	301	GTGACGGGGAAAAGTGAATCCGGAGAGAACGAATCAAGGTGGTGTGTTAGAACATATG	360
Db	301	GTGACGGGGAAAAGTGAATCCGGAGAGAACGAATCAAGGTGGTGTGTTAGAACATATG	360
Qy	361	AAAGTGGAGAGATTTTCCTCGAAGGTTCCAGACACGGCAGCTCAACCGTGGAGTTTWT	420
Db	361	AAAGTGGAGAGATTTTCCTCGAAGGTTCCAGACACGGCAGCTCAACCGTGGAGTTTWT	420
Qy	421	GGAAGTTTCCACTCGCAAACTTTCGACTTCCCTCGAGGTGGAATCATAGGCTGT	480
Db	421	GGAAGTTTCCACTCGCAAACTTTCGACTTCCCTCGAGGTGGAATCATAGGCTGT	480
Qy	481	CTGATAGAGTTTCAACAGACACCGCAGAGATCTCGACATCTGGGTGGAACAAGATGCT	540
Db	481	CTGATAGAGTTTCAACAGACACCGCAGAGATCTCGACATCTGGGTGGAACAAGATGCT	540
Qy	541	GAAACCGAGAGAAACTTTGGAAAAGTGAAGTGAATGAAGTCTTCAGAAAGCCGGTG	600
Db	541	GAAACCGAGAGAAACTTTGGAAAAGTGAAGTGAAGTCTTCAGAAAGCCGGTG	600
Qy	601	GGACAGGAGATGACGATCAAACTTGGAGAGGAAAGAAAAGATTGAAACATCCAAACGA	660
Db	601	GGACAGGAGATGACGATCAAACTTGGAGAGGAAAGAAAAGATTGAAACATCCAAACGA	660
Qy	661	TTCCGTCGAAGGGGAGTTTCACTCTCGAAAACCGCAGAGTTCTGGAGGCTCGAAGATCATAT	720
Db	661	TTCCGTCGAAGGGGAGTTTCACTCTCGAAAACCGCAGAGTTCTGGAGGCTCGAAGATCATAT	720
Qy	721	CTTTATCTCTTCAAGGTGGAACCTTGAAAAAGACGAGTACACTCTGGAACATCGGAATCAGA	780
Db	721	CTTTATCTCTTCAAGGTGGAACCTTGAAAAAGACGAGTACACTCTGGAACATCGGAATCAGA	780
Qy	781	ACGATCAAGTGGAGCGAAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGCG	840
Db	781	ACGATCAAGTGGAGCGAAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGCG	840
Qy	841	TTTGGAAAAGCAGAGAAATCCCGGTTCTGGGGCAGGGGCACTTTATCCATGATGATG	900
Db	841	TTTGGAAAAGCAGAGAAATCCCGGTTCTGGGGCAGGGGCACTTTATCCATGATGATG	900
Qy	901	AAAGACTTCAACCTTCTGAGTGTATCAACCGGAAATCTTTCAGAGACCTTCATCATCT	960
Db	901	AAAGACTTCAACCTTCTGAGTGTATCAACCGGAAATCTTTCAGAGACCTTCATCATCT	960
Qy	961	TACAGTGAAGGTGGCTGGATCTTCCGACAGACTCGGAATCTTGTGATGAGCAAAACC	1020
Db	961	TACAGTGAAGGTGGCTGGATCTTCCGACAGACTCGGAATCTTGTGATGAGCAAAACC	1020
Qy	1021	CCGACGTTGGATATCAAAAGTACCACTATCCCGAGACTCAAAAGATGACAAAGAC	1080
Db	1021	CCGACGTTGGATATCAAAAGTACCACTATCCCGAGACTCAAAAGATGACAAAGAC	1080
Qy	1081	AACATGAAGAGATGATGACAGACAAAGAACATCCCAAGTGTATCATGTGAGGTGTG	1140
Db	1081	AACATGAAGAGATGATGACAGACAAAGAACATCCCAAGTGTATCATGTGAGGTGTG	1140
Qy	1141	GCGAAGCAACCAAGTCCAACTCCAGACGCGGAGGGTTTCTTCAAGCCCTTATGAG	1200
Db	1141	GCGAAGCAACCAAGTCCAACTCCAGACGCGGAGGGTTTCTTCAAGCCCTTATGAG	1200

Db	1141	GGAAACGACAGAGTCCAAACATTCACAGACGCGGAGGGTTTTCTTCAAAAGCCCTTTATGAG	1200
Qy	1201	ACTGCCAATGAATGAGATCGAACACGCCCGGTTGTACTGGTAGCATGATGACGCACCA	1260
Db	1201	ACTCCCAATGAAATGAGATCGAACACGCCCGGTTGTACTGGTAGCATGATGACGCACCA	1260
Qy	1261	GACGAGGAACAABAAGACGTGGCGGCTGAAGTACTTCGACATCGCTGTGTGAACAGATAC	1320
Db	1261	GACGAGGAACAABAAGACGTGGCGGCTGAAGTACTTCGACATCGCTGTGTGAACAGATAC	1320
Qy	1321	TACGCGTGTGATCTATCAGGGAAAGATAGAAAGAAGACTTCAAGCTCTGSAAAAAGAC	1380
Db	1321	TACGCGTGTGATCTATCAGGGAAAGATAGAAAGAAGACTTCAAGCTCTGSAAAAAGAC	1380
Qy	1381	ATTAGAAGAGCTCTATGCAAGGCACAGAAAGCCACTCTTGTCAAGAAATTGGTGGCGAC	1440
Db	1381	ATTAGAAGAGCTCTATGCAAGGCACAGAAAGCCACTCTTGTCAAGAAATTGGTGGCGAC	1440
Qy	1441	GCGATAGCTGGCATCTCCATACGATCCACCTTCAATGTTCTCCGAAGGTACCAAGCAGAG	1500
Db	1441	GCGATAGCTGGCATCTCCATACGATCCACCTTCAATGTTCTCCGAAGGTACCAAGCAGAG	1500
Qy	1501	CTCGGTGAAAAGAGATCAGGCTCCCTTTTGAAAAAAGACTACATCATCGGAACACAGTG	1560
Db	1501	CTCGGTGAAAAGAGATCAGGCTCCCTTTTGAAAAAAGACTACATCATCGGAACACAGTG	1560
Qy	1561	TGGGCTTTTGAGATTTTAAAGACTCTCAGAAATGTGAAGAAGCCATTCTCAACCAAG	1620
Db	1561	TGGGCTTTTGAGATTTTAAAGACTCTCAGAAATGTGAAGAAGCCATTCTCAACCAAG	1620
Qy	1621	GGTGTGTTTACAGAAGACAGCAAACTCGTGTCTCATGTACGTAGAAGACTGTGG	1680
Db	1621	GGTGTGTTTACAGAAGACAGCAAACTCGTGTCTCATGTACGTAGAAGACTGTGG	1680
Qy	1681	AGTAGAGTTT	1689
Db	1681	AGTAGAGTTT	1689

```

RESULT 3
US-10-136-841-5
; Sequence 5, Application US/10136841
; Publication No. US20030082176A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan
; APPLICANT: Beverley, Stephen
; TITLE OF INVENTION: SUBCELLULAR TARGETING OF THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-007
; CURRENT APPLICATION NUMBER: US/10/136,841
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US 60/287,531
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/304,609
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/329,461
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/351,276
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A recombinant sequence incorporating a signal peptide sequence,
; OTHER INFORMATION: the mature human beta-glucuronidase sequence, a bridge of three
; OTHER INFORMATION: amino acids, and an IGF-II sequence
; NAME/KEY: CDS
; LOCATION: (1)..(2166)
US-10-136-841-5

```

Query Match 10.3%; Score 174; DB 14; Length 2169;
 Best Local Similarity 52.2%; Pred. No. 1.2e-44;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGAAAGAGTACCTCTGAGATCGAATCAGAACGATCAGCTGGAGC 796
 DB 941 TGGGGCTGTGTGCTTCTTACACACTCCCTGTGGGATCCGCACTGTGGCTGACCA 1000
 QY 797 AAGAAAGCTCTATCTGAAACGGGAACTGTCTTTTGAAGGGCTTTGAAAGCAGAG 856
 DB 1001 AAGAGCAATCTCTCAATAGGAAACCTTTCTATTCCACGGTCTCAAGCATGAGG 1060
 QY 857 AATTCCTCTTGGGGCAGGGCACCCTTTATCATGATGATAAAGCTTCAACCTTC 916
 DB 1061 ATGGGAGATCCGAGGAAAGGGCTTGCATGCGCTGCTGTGTGAAGATTTCAACTGC 1120
 QY 917 TGAAGTGAATCAACCGAATTTCTTCAAGACTCTCACTATCTTCAAGTGAAGTGC 976
 DB 1121 TTGCTGTGTGTGATCCAAAGCTTTCCTTCCGATCAGCACTACCTTATGAGAGTGA 1180
 QY 977 TGGATCTTGGCGAGAGCTCGGATCTTGTGATGATGAAGAACCCCGACCTTGTATCA 1036
 DB 1181 TGCATATGTGTACCGCTATGAGATGTGTGTATGATGATGATCCGCGGTGGCTG 1240
 QY 1037 CAAGGTACCA--CTACATCCCGAGACTCAGAAAGATAGCAAGACAATTAAGAA 1093
 DB 1241 CGCTCCGCAAGTTCTTCAACAACGTTTCTGTGATCACCACATGAGGTATGAGAG 1300
 QY 1094 TGATGAGACAGACAGAAACCATCCCACTGTGATCATGTGAGTGTGGACAGAACAG 1153
 DB 1301 TGTGTGTAAGAGACAAAGACCAACCCGCGGTGTGATGTGTGTGTGTGTGTGTGTGT 1360
 QY 1154 AGTCCAACTATCCAGACCGGAGGGTTCTTCAAAAGCCCTTATGAGACTGSCAATGAA 1213
 DB 1361 CGTCCACTTGAATCTGT 1420
 QY 1214 TGTGATGAAACAGCCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1273
 DB 1421 TGGACCCCTCCCGCTGT 1468
 QY 1274 GAGAGTGGGGCTGAAGTACTTGCACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1333
 DB 1469 CAGACAAAGGGGCTCCGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1528
 QY 1334 TCTATCAGGAAAGATAGAAAGAGACTTCAAGCTGTGAAAGAAAGATAGAGCTCT 1393
 DB 1529 ACGACTAGGGGCACTGAGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1588
 QY 1394 ATGCAAGGACAGAAAGCCCATCTTTGTACAGAAATTCGTGTGGAGCGCATAGTGA 1453
 DB 1589 ATAAAGATATCAGAAAGCCCATTTATCAGAGCGATGTGAGAGCAAAACGATTCAG 1648
 QY 1454 TCCACTAGATCCACTCAAAATGTTCTCCGAAGTCCAGAGCAAGCACTGTTGAAAG- 1512
 DB 1649 TTCAACAGATCCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1708
 QY 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCTGAAACACAGTGTGGCT 1567
 DB 1709 ACCATCTGGGCTGTGATCAAAAAAGCAAGAAATATGTGTGTGTGTGTGTGTGTGT 1768
 QY 1568 TTGCAATTTTAAAGCTCTCAAGATGTGAAAGAACCATTTCTCAACCAAGAGGTGT 1627
 DB 1769 TTGCGAATTTATATATGAAACAGTACCGAGAGAGTGTGGGAAATTAAGAGGGATCT 1828
 QY 1628 TCAAGAGACAGACCAACCACTGTTGTCTCATGTACTGAG 1671
 DB 1829 TCACTGGGAGAGCAACCAAAAGTCAAGCTTCTTTTGGCA 1872

RESULT 4
 US-10-272-531A-5
 ; Sequence 5, Application US/10272531A
 ; Publication No. US2004005309A1

GENERAL INFORMATION:
 ; APPLICANT: Lebowitz, Jonathan H
 ; APPLICANT: Beverly, Stephen
 ; APPLICANT: Sly, William S.
 ; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
 ; FILE REFERENCE: SYM-009
 ; CURRENT APPLICATION NUMBER: US/10/272,531A
 ; PRIOR FILING DATE: 2002-10-16
 ; PRIOR APPLICATION NUMBER: US 60/384,452
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: US 60/386,019
 ; PRIOR FILING DATE: 2002-06-05
 ; PRIOR APPLICATION NUMBER: US 60/408,816
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 2169
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: A recombinant DNA sequence incorporating a signal peptide sequenc
 ; OTHER INFORMATION: e, the mature human beta-glucuronidase sequence, a bridge of thre
 ; OTHER INFORMATION: e amino acids, and an IGF-II sequence
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(2166)
 ; OTHER INFORMATION:
 ; US-10-272-531A-5

Query Match 10.3%; Score 174; DB 16; Length 2169;
 Best Local Similarity 52.2%; Pred. No. 1.2e-44;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGAAAGAGTACCTCTGAGATCGAATCAGAACGATCAGCTGGAGC 796
 DB 941 TGGGGCTGTGTGCTTCTTACACACTCCCTGTGGGATCCGCACTGTGGCTGACCA 1000
 QY 797 AAGAAAGCTCTATCTGAAACGGGAACTGTCTTTTGAAGGGCTTTGAAAGCAGAG 856
 DB 1001 AAGAGCAATCTCTCAATAGGAAACCTTTCTATTCCACGGTGTCAAGCATGAGG 1060
 QY 857 AATTCCTCTTGGGGCAGGGCACCCTTTATCATGATGATGAAGAAAGCTTCAACCTTC 916
 DB 1061 ATGGGAGATCCGAGGAAAGGGCTTGCATGCGCTGCTGTGTGAAGATTTCAACTGC 1120
 QY 917 TGAATGATCAACCGGAATTTCTTCAAGACTCTTCACTATCTTACATGAAAGTGC 976
 DB 1121 TTGCTGT 1180
 QY 977 TGGATTTCCGAGAGCTCGGAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1036
 DB 1181 TGCATGTGTGTACCGCTATGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1240
 QY 1037 CAAGGTACCA--CTACATCCCGAGACTCAGAAAGATAGCAAGACAATTAAGAA 1093
 DB 1241 CGTCCGAGAGTTCTTCAACAACGTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1300
 QY 1094 TGAATGACAGACAGAAACCATCCAGTGTGATGATGAGTGTGGAGCGAACAG 1153
 DB 1301 TGT 1360
 QY 1154 AGTCCAACTATCCAGAGCGGAGTTTCTTCAAGCCCTTATGAGACTGCAATGAA 1213
 DB 1361 CGTCCACTTGAATCTGT 1420
 QY 1214 TGGATGAAACAGCCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1273
 DB 1421 TGGACCCCTCCCGCTGT 1468
 QY 1274 GAGAGTGGCGCTGAAGTACTTGCACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1333
 DB 1469 CAGACAAAGGGGCTCCGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1528


```

; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; TITLE OF INVENTION: TREATMENT OF LYSOSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/10/421,175
; CURRENT FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: US/09/715,858
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patemtein Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1956)
; US-10-421-175-1

```

```

Query Match      10.2% Score 172.4; DB 15; Length 1956;
Best Local Similarity 52.1%; Pred. No. 3.6e-44;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

```

Qy 737 TGGAACTTGAAMAAAGAGTACCTCTGACATCGAATGAGAACGATCAGCTGGAGCG 796
Db 938 TGGGGCTGTGTGCTGACTTCTACACACTCCCTGTGGGATCCGACCTGTGCTGCACCA 997
Qy 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
Db 998 AGAGCAGTCTCTCATCATGTGGAAACCTTTCTATTTCCAGGGTGTCAACAAGATGAGG 1057
Qy 857 AATTCCTGCTCTGGGAGGAGGACCTTTTATTCATGATGATTAAGAAGCTTCAACCTTC 916
Db 1058 ATGGGAGATCGAGGAGGAGGCTTGCAGCTGGCCCTGTCTGTGAAGACTTCAACCTTC 1117
Qy 917 TGAAGTGAATCAACCGCAATTTCTTCAAGACCTTCACTATCTTCAAGTGAAGAGTGC 976
Db 1118 TTCCGTGCTGTGTGTCACCAAGCTTTCGTTACAGCACTACCTCTATGACAGAGAGTGA 1177
Qy 977 TGGATCTTGGCCGACGACTCGGAATCTTGTGATGAGCAAGCCCGCAGCTTGTATCA 1036
Db 1178 TGCAGATGTGTGACCGCTATGAGATTTGTGATGATGATGATGATGATGATGATGATGAT 1237
Qy 1037 CAAGGTACCA---CTACATCCGAGACTCAGAAATAGCAGAAAGCAACATAAGAGAA 1093
Db 1238 CGTGCAGCAGTCTTCAACAACGTTTCTGCAATCAACATGCAAGTATGAGAAAG 1297
Qy 1094 TGATCGACAGACAGAAACCAATCCCACTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
Db 1298 TGGTGCCTGAGGAGAAAGAACCCCGGATGTGATGTGTGTGTGTGTGTGTGTGTGTGTGT 1357
Qy 1154 AGTCCAAACATCGACGCGGAGGATTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
Db 1358 CGTCCCACTGATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1417
Qy 1214 TGGATCGAACAAGCCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1273
Db 1418 TGGACCCCTCCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1465
Qy 1274 GAGAGTGGGCTGAGATCTTGAACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1333
Db 1466 CAGCAGAGGGGCTTCCTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1525
Qy 1334 TCTATCAGGAGAGATGAGAAAGCACTTGAAGCTTGAAGAAAGACATGAGAGAGCTCT 1393
Db 1526 ACGACTACGGGCACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1585
Qy 1394 ATGCAAGGACAGAAAGCCATCTTTGTCAACAAGATTCGGTGGGAGCGCAGTACCTGGCA 1453
Db 1586 ATAGAAGTATCAGAAACCCATTAATTCAGAGCGATGTGAGAGCAAAAGATTCAGAGGGT 1645
Qy 1454 TCACACTGATCCACTCAATGTCTCGAAGAGTACCAAGAGAGCTGTGTGAAG- 1512

```

```

Db 1646 TTCACGAGATCCACCTGTGATGTCACTGAGAGTACCAAGAAAGTCTGAGAGAGT 1705
Qy 1513 ACAGTACAGCTCTTTTGAAGAA---AGACTACATCATCGGAACAGAGTGGGCT 1567
Db 1706 ACATCTGGGTCTGTGATCAAAAACGAGAAATATGTGTGTGAGAGCTTATTTGAAAT 1765
Qy 1568 TTGCAGATTTTAAAGTCTCTGAGATGTGAGAAAGCCATTTCAACCAAGAGGTGT 1627
Db 1766 TTGGCGATTTTCAATGACTGAACAGTCAACCGACGAGAGTGTGGGAAATTAAGAGGATCT 1825
Qy 1628 TCACAAAGACAGACAAACCACTGTTGCTCATGTACTGAGA 1671
Db 1826 TCATCGGCGAGAGCAACCAAAAGTACGAGCTTCTTTGCGA 1869

```

RESULT 7

```

US-10-388-360-343
Sequence 343, Application US/10388360
Publication No. US20030225528A1

```

```

; GENERAL INFORMATION:
; APPLICANT: GENOMIC HEALTH
; APPLICANT: Baker, Joëlle B.
; APPLICANT: Cronin, Maureen T.
; APPLICANT: Kiefer, Michael C.
; APPLICANT: Shak, Steve
; APPLICANT: Walker, Michael Graham
; TITLE OF INVENTION: GENE EXPRESSION PROFILING IN BIOPSIED TUMOR TISSUES
; FILE REFERENCE: 39740-000105
; CURRENT APPLICATION NUMBER: US/10/388,360
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: US 60/412,049
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/364,890
; PRIOR FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 343
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-388-360-343

```

```

Query Match      10.2% Score 172.4; DB 15; Length 2191;
Best Local Similarity 52.1%; Pred. No. 3.8e-44;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

```

Qy 737 TGGAACTTGAAMAAAGAGTACCTCTGACATCGAATGAGAACGATCAGCTGGAGCG 796
Db 964 TGGGGCTGTGTGCTGACTTCTACACACTCCCTGTGGGATCCGACCTGTGCTGCACCA 1023
Qy 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
Db 1024 AGAGCAGTCTCTCATCATGTGGAAACCTTTCTATTTCCACGATGTCAACAAGCATGAGG 1083
Qy 857 AATTCCTGCTTGGGAGGAGGACCTTTTATCAATGATGATTAAGAAGCTTCAACCTTC 916
Db 1084 ATGGGAGATTCGAGAGGAAAGGCTTCACTGGCCCTGTGTGAAGACTTCAACCTTC 1143
Qy 917 TGAAGTGAATCAACCGCAATTTCTTCAAGACCTTCACTATCTTAAAGTGAAGAGTGC 976
Db 1144 TTCCGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1203
Qy 977 TGGATCTTGGCCGACGACTCGGAATCTTGTGATGAGAAAGCCCGCAGCTTGTGTATCA 1036
Db 1204 TGCAGATGTGTGACCGCTATGAGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1263
Qy 1037 CAAGGTACCA---CTACATCCGAGACTCAGAAATAGCAGAAAGCAACATAAGAGAA 1093
Db 1264 CGTCCCGAGATTTCTTCAACAACGTTTCTGTGATCAACAATGAGTATGAGAAAG 1323
Qy 1094 TGATCGACAGACAGAAACCAATCCCACTGTGATCATGTGAGTGTGGAGCAAGCAACAG 1153
Db 1324 TGGTGCCTGAGGAGCAAGAAACCCCGGATGTGATGTGTGTGTGTGTGTGTGTGTGTGTGT 1383

```

QY 1154 AGTCAACATCCAGACGGGGGTTCTTCAAGCCTTTATGAGACTGCCAATGAAA 1213
DB 1384 GGTCCACCTAGAAATCTGTGGCTACTTGAAGATGGTATCTCTCACCAAAATCTT 1443
QY 1214 TGGATCGAACACGCCCCGTTGTCAATGATGAGCATATGAGACGACACAGAGAAACAA 1273
DB 1444 TGGACCCCTCCCGGCTGTGACCTTTGTAGCACTTAACTATGACG----- 1491
QY 1274 GAGACGTGGCGCTGAAGTACTTGCATCTGTGTGTGAACAGTACTACGGCTGTACA 1333
DB 1492 CAGACAAAGGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGTACTACTTGTGTATC 1551
QY 1334 TCTATCAGGGAAGATATGAAAGAGACTTCAAGCTCTGGAATAAAGATAGAGCTCT 1393
DB 1552 ACGATTACGGGACCTGTGAGTTGATTCAGCTGACCTGCGCACCGATTGGAATCGGT 1611
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGTACAGAAATTCGGTCCGAGCGGATAGCTGGA 1453
DB 1612 ATTAAGATATACAGAAAGCCCATTTATCAGAGGAGATATGAGACAGAAAGATTGCAAGGT 1671
QY 1454 TCCACTAGATCCACTCAATGTTCTCCGAAGATCCAGACAGAGCTGTTGAAAAG- 1512
DB 1672 TTCAACAGGATCCACTGTGATGTTCACTGAAGAGTACAGAAAGTCTGTAGAGCAGT 1731
QY 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAACACAGCTGTGGGCT 1567
DB 1732 ACCATCTGGGTCTGGATCAAAAACGAGAAATATGTGGTGGAGAGCTCATTTTGAATTT 1791
QY 1568 TTGCAGATTTTAAAGCTCCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAGAGGTGTTT 1627
DB 1792 TTGCGGATTTTCACTGACTGAACAGTCAACGAGAGAGTGTGGGGAATTAAGAGGAGATCT 1851
QY 1628 TCACAAGAGACAGACCAACCAATCTGTTGCTCATGTACTGAGA 1671
DB 1852 TCACCTGGCAGAGACCAAAAAGTGACGCGTCTTTTGGCA 1895

RESULT 8
US-10-429-802-16
; Sequence 16, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: WONG, KA YIN
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-16

Query Match 10.2%; Score 172.4; DB 15; Length 2191;
Best Local Similarity 52.1%; Pred. No. 3.8e-44;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGAAGCTTGAAGAAAGAGTACTCTGACATCGAATCAGATCAGACTGGAGCG 796
DB 964 TGGGGCCGTGTCTGACTTTTACACATCCCTGTGGGATCCGCACTGTGCTGTACCA 1023
QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTTGAAGACAGAGG 856
DB 1024 AGAGCAAGTCTCTCATCATGTGGAACCTTTCTATTTCCAGGCTGTCAACAGATGAGG 1083
QY 857 AATTCGCCGTTCTGGGGAGGGCACCTTTTATCCATGTATGATTAAGACTTCAACCTTC 916

DB 1084 ATGCGGATATCCGAGGGAAGGGCTTCACTGGCCGCTCTGTGTGAAGACTTCAACTTC 1143
QY 917 TGAAGTGAATCAACGCAATTTCTTTCAGACCTCTCACTATCTTACATGAAAGTGGC 976
DB 1144 TTGCTGTGCTGTGTGTCCAAAGCTTTTCGTACAGGCACTAACCCCTATGACAGGAATGTA 1203
QY 977 TGGATCTTGGCGACAGACTCGGAATCTTGTGTATGAGAAAGCCCCGACCTGTGTATCA 1036
DB 1204 TGCAGATGTGTGACCGCTATGTGAGATTTGTGTATGAGAGTGTCCGCGGTGGCTTGG 1263
QY 1037 CAAGGTACCA---CTACAAATCCCGAGCTCAGAAATGACAGAAAGACAAATTAAGAGTA 1093
DB 1264 CCGTCCCGCAGTTCTTCAACCAAGCTTTCTCTGATACCAACATGAGATGATGAGAAAG 1323
QY 1094 TGAATGACAGACACAGAAACCAATCCAGTGTATCATGTGAGTGTGCGGAAAGCAACGAG 1153
DB 1324 TGTGCTGTAGGGAACAAAGAACCAACCCCGGCTGTGATGTGTGTGTGCTGACAGAGCTTG 1383
QY 1154 AGTCAACATTCAGACCGGAGGGTTCTTCAAAAGCCCTTTATGAGACTGCGCAATGAAA 1213
DB 1384 GGTCCACCTTGAATCTGCTGCTACTAATTGAAGATGTGATGCTCAACCAATCTCT 1443
QY 1214 TGGATCGAACACGCCCCGTTGTCAATGTGAGAGATGATGAGACGACACAGAGAGAACAA 1273
DB 1444 TGAACCCCTCCCGGCTGTGACCTTTGTGAGCACTTAACTATGACG----- 1491
QY 1274 GAGAGTGGCGCTGAAAGTACTTGCATCTGTGTGTGAACAGGATCTTACAGGCTGTGTA 1333
DB 1492 CAGACAAAGGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTTGTGTATC 1551
QY 1334 TCTATCAGGGAAGATATGAAAGAGACTTCAAGCTCTTGAAAAGAAACATAGAGAGCTCT 1393
DB 1552 ACGATCAGGCAACCTGTGAGTTGATTCAGCTGACAGCTGCCACCGATTGGAATCGT 1611
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGTACAGAAATTCGGTCCGAGCGGATAGCTGGA 1453
DB 1612 ATTAAGATATCAGAAAGCCCATTTATTCAGAGGAGATATGAGACAGAAAGCATTTGAGGCT 1671
QY 1454 TCCACTAGATCCACTCAATGTTCTCCGAAGATCCAGACAGAGCTCGTTGAAAAG- 1512
DB 1672 TTCAACAGGATCCACTGTGATGTTCACTGAAGAGTACAGAAAGTCTGTAGAGCAGT 1731
QY 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATGTGGAACACAGCTGTGGGCT 1567
DB 1732 ACCATCTGGGTCTGGATCAAAAACGAGAAATATGTGTGGAGAGCTCATTTTGAATTT 1791
QY 1568 TTGCAGATTTTAAAGCTCCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAGAGGTGTTT 1627
DB 1792 TTGCGGATTTTCACTGACTGAACAGTCAACGAGAGAGTGTGGGGAATTAAGAGGAGATCT 1851
QY 1628 TCACAAGAGACAGACCAACCAATCTGTTGCTCATGTACTGAGA 1671
DB 1852 TCACCTGGCAGAGACCAAAAAGTGACGCGTCTTTTGGCA 1895

RESULT 9
US-10-430-503-7
; Sequence 7, Application US/10430503
; Publication No. US20040005684A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: LAN, KENG-LI
; APPLICANT: OU-YANG, FU
; APPLICANT: LIU, JAM-CHING
; APPLICANT: LAN, KENG-HSIN
; TITLE OF INVENTION: TARGETING PROTEINS TO DELIVER THERAPEUTIC OR DIAGNOSTIC
; FILE REFERENCE: UTSC:797US
; CURRENT APPLICATION NUMBER: US/10/430,503
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: 60/383,063
; PRIOR FILING DATE: 2002-05-06

Db 1612 ATAGAGATATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGACAGAAACGATTCAGAGCT 1671
Qy 1454 TCACCTACGATTCACCTCAATATGTTCTCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG- 1512
Db 1672 TTCACACGAGATTCACCTCTGATGTTCACTGAAAGATACAGAAAAGTCTGCTAGAGCAGT 1731
Qy 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAAACACACGTGTGGGCT 1567
Db 1732 ACCATCTGGCTCTGGATCAAAAACGAGAAAATATGTGGTTGGAGAGCTCATTTTGAATTT 1791
Qy 1568 TTGCGATTTTAAAGCTCCTCAGATGTAGAAAGACCCATTCTCAACCAAGAGGTGTTT 1627
Db 1792 TTGCGGATTTTCACTGACTGAACGATACCGACGAGAGTCTGGGGAAATAAAAAGGGGATCT 1851
Qy 1628 TCACAGAGACAGACACCCAAACTGTTGCTCATGTACTGAGA 1671
Db 1852 TCACCTGGCAGAGACACCAAAAAGTGACGCTTCTTTTGGCA 1895

RESULT 11
US-10-388-934-225
; Sequence 225, Application US/10388934
; Publication No. US20040005547A1
; GENERAL INFORMATION:
; APPLICANT: Boeas, Franziska
; APPLICANT: Suter-Dick, Laura
; APPLICANT: Wolf, Detlef
; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
; FILE REFERENCE: 21199
; CURRENT APPLICATION NUMBER: US/10/388,934
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 02005336.9
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 02015657.6
; PRIOR FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 862
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 225
; LENGTH: 2472
; TYPE: DNA
; ORGANISM: Rattus norvegicus (No. US20040005547A1way rat)
US-10-388-934-225

Query Match 9.9%; Score 166.8; DB 16; Length 2472;
Best Local Similarity 51.3%; Pred. No. 2.7e-42;
Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;

Qy 751 GACGAGTACCTCTGAGATCGGAATTCAGAAAGATCAGCTGAGAGAGAAAGGCTCTAT 810
Db 946 GACTTCTATACCTCTCTGTCGGGATTCGACAGTGGCTGTCAAAAAGAGCAAGTTCTCTC 1005
Qy 811 CTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAGACAGAGAAATCCCGTTCG 870
Db 1006 ATAAATGGGAACCTTTTACTTCCAAAGCGTCACAAAGCATGAGATTCAGATATCCGA 1065
Qy 871 GGGCAGGGCACCCTTTATTCATTGATGATTAAGAAGCTTCAACCTTCTGAAGTGAAC 930
Db 1066 GGGAGAGGCTTCGATGGCTCTGCTGATTAAGAAGATTCACCTCTCCGTTGGCTCGGG 1125
Qy 931 GCGAATCTTTCAGAGCCTCTCACTATCTTAAGTGAAGAGTGGCTGATCTTCCGAC 990
Db 1126 GCAAAATCTCTTTCGACAGCCACTATCTCTGAGAGAGTACTTCAGCTCTGTGAC 1185
Qy 991 AGACTCGGAATCTTGTATAGACGAAGCCCGCAGTGTGATCAAGAGTACGA---C 1047
Db 1186 CGATATGGAATTTGTGCTCATGATAGTGTCCCGGTGTGGGATGTGCTGCCAGAGT 1245
Qy 1048 TACAATCCCGAAGTCAAGAGATGACAGAAACCATTAAGAAATATGACAGACAG 1107
Db 1246 TTGGCAACGTGTCTCTTGGGACCACTAGAGGTGATGAGAGAGCTGTGTGGGAG 1305
Qy 1108 AAGAACATCCCAAGTGTATCATGTGAGGTGTGGGAAACGAAGTCCCAACATCA 1167

Db 1306 AAAAATCACCTCGCGGTGTATGTGTCTGTGGCAATGAGCTCTTCTTTCTGAAA 1365
Qy 1168 GACGGGAGGGTTTCTTCAAGCCCTTTATGAGACTGCAATGAATGATGAAACAGC 1227
Db 1366 COTGGCGGATTTTACTTCAAGACGCTGATTCGCCACACCAAAAGCCCTGAGCCCACT 1425
Qy 1228 CCCGTTGATGCTGAGATATGAGACGACACAGACAGAAACAGAGACGTGGCGCTG 1287
Db 1426 CCGTGACCTTTGTGAGCAATA-----CCAGATATGACGAGACATGGGGCC 1473
Qy 1288 AAGTACTTGAATCTGTGTGTAAGAGTACTACGCGTGTATCATCTATCAGGGAAG 1347
Db 1474 CCGTACGTGACGCTGATTTGTGTAAACGTTACTTATCTGTATCATGACCTAGCGGAT 1533
Qy 1348 ATAGAAAGAGACTTCAAGCTCTGAAAAAAGACATAGAAAGCTTATGACAGACAGA 1407
Db 1534 CTGAGAGTGAATTCAGCTGACAGCTGACCTGACAGATTGAGAACTGGTATAGATGACAG 1593
Qy 1408 AAGCCATCTTTGTCAAGAAATTTGGTGGACCGGATAGCTGGCATCATGACATCA 1467
Db 1594 AAGCCAAATTTATCCAGAGAGATAGAGACACACCGCTCTGGGGCTTCATGAGGATCA 1653
Qy 1468 CCGTAAATGTTCTCGAAGAGTACCAAGCAGAGCTCGTTGAAAAGAGATCAGGCTCTT 1527
Db 1654 CCGCATGTTCACTGAGAGATACAGACACCTCTCTGAGAAATTTATCATTTGATCTG 1713
Qy 1528 TTGAA-----AAAAGTACATCATCGGAACACACGTTGGGCTTTGAGATTTTAA 1581
Db 1714 GATGAGAAAGAGAAAGATATGTATCGGAAGGCTCATCTGGAATTTTGGTCACTTCA 1773
Qy 1582 ACTCTCAGAAATGAGAGAGCCATTTCAACCAAGAGGTGTTTCAAGAGACAGA 1641
Db 1774 ACGAACAGTACCACTGAGAGTAAACAGAAACAAAGAGGAGATCTTCACTCGACAGA 1833
Qy 1642 CAACCAAACTGTTGCTCATGTACTGAGA 1671
Db 1834 AAGCCAAAGATGAGAGCTTCAATTTTGGCA 1863

RESULT 12
US-10-421-175-3
; Sequence 3, Application US/10421175
; Publication No. US2003021941A1
; GENERAL INFORMATION:
; APPLICANT: Podaskoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VARIANTS FOR THE
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/10/421,175
; PRIOR FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: US/09/715,858
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1947)
US-10-421-175-3

Query Match 9.4%; Score 159.6; DB 15; Length 1947;
Best Local Similarity 50.6%; Pred. No. 5e-40;
Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

Qy 739 GAACCTGAAAAGACGATCACTGTGACATCTGGAATTCGAAGATAGCTGGAGAG 798

Db 928 GAGTCTGTGACTGACTACTACTACCCCTTCCTATCGGGATTGAAACAGTGGCTGTCAACAAG 987
 Qy 799 AAGAGGCTCTATCTGTAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGAA 858
 Db 988 AGCAATCTCTCATTAACGGGAAAGCCCTTCTATTTTCAAGGGGTCAATTAAGCAGAGAT 1047
 Qy 859 TTCCCGCTTCTGGGGGAGGAGCACTTTTATCCATTGATGATTAAGAACTTCAACCTTCG 918
 Db 1048 TCAGATATCCGAGGGAAAGGCTTCCAGCTGGCCGCTCTGTGTAAGAGATTTCACCTGCTC 1107
 Qy 919 AAGTGATCAAGCCGAAATCTTTCAAGGACCTCTCACTATCTTACAGTAAGAGTGGCTG 978
 Db 1108 CATTGCTGTGGGGCAAAATCTTTGTGTACAGCCCACTATCCCTACTCAGAGAGGACTCTT 1167
 Qy 979 GATCTTGGCAGACGACTGGAAATCCCTGTGATAGCAAGAGCCGCCAGCTGG--TATC 1035
 Db 1168 CAGCTCTGTGACCGATACGGGATTTGTGGTATCATGATGTGTCCGGTGTGGCAATTTG 1227
 Qy 1036 ACAAGGTACCACTACATCCCGAGACTCGAAGATAGCAGAAACATTAAGAAATG 1095
 Db 1228 CTACTCAGAGTTTGGCAACAGTCACTTGGGACACACCTAGAGTATGAGAGAGCTG 1287
 Qy 1096 ATCGACAGACAAAGAACATCCAGTGTGATCATGTGAGTGTGGCAACGAAACGAG 1155
 Db 1288 GTTCGCCGGGACAAAAATCACCTCGGTTGTGATGTGTGTGTGGCCAAATGAGCTTCC 1347
 Qy 1156 TCCAAACATCCAGACGGGAGGTTCTTCAAGGCCCTTTATGAGACTGCGCAATGAATG 1215
 Db 1348 TGTGCTGTGAACCCGCGCATATTACTTTAAGAGCTGATACCCACACCAAGAGCCCTG 1407
 Qy 1216 GATCGAACACGCCCGCTGTGTCTGTGATGATGAGCAGCAGCAGAGAGAAACAAGA 1275
 Db 1408 GACCTCACCCGCTCCGTGACCTTTGTGAGCA-----ACGCCAAATATGATGCA 1455
 Qy 1276 GACGTGGGCTGAAATCTGTGACATCGTGTGTGGAACAGTACTAGCGCTGTATATC 1335
 Db 1456 GACCTGGGGGCCCCGTACGTGATGTATCTGTGAACAGCTACTTTCTTGTATATCAT 1515
 Qy 1336 TATCAGGGAAGATTAAGAAAGACTTCAGCTCTGAAAAAAGCATTAAGAGACTCTAT 1395
 Db 1516 GACTATGGGCAATTTGAGAGTGTATCAAGCAGCTGAATACCAAGTTTGAAGACTGGTAT 1575
 Qy 1396 GCAAGGACAAAGAACCATCTTGTCTACAGAAATTCGTGGGACGGGATAGCTGATC 1455
 Db 1576 AAGAGCATCAGAAAGCCATTAATCCAGGCGAGATGAGCAGACGCAATCCCGAGATC 1635
 Qy 1456 CACTACGATCCACCTCAATGTCTCCGAGAGTACCAAGCAGAGCTCGTTGAAGAAAGC 1515
 Db 1636 CACGAGGACCGGCTCGCATTTCAAGAGAGATACCAAGAGCTGTCTGTGAGAAATTA 1695
 Qy 1516 ATCAGGCTCTTTTGA--AAAGACTACATATCGGAACACAGTGTGGGCTTT 1569
 Db 1696 CATTCAATTTCTGATCAGAAACGTAAAGATACTGTGCGAGCTCATCTGGAATTT 1755
 Qy 1570 GCAGATTTTAAAGCTCCGAGATGTGAGAAAGCCATTCACCAACCAAGGCTTTTC 1629
 Db 1756 GCCAATTTCAAGAACCAAGTACCGCTGAGATTAATCGGAACCAAGAGGAGATCTTC 1815
 Qy 1630 ACAAGAGACAGAACCAAACTCGTGTCTCATGACTACTGAGA 1671
 Db 1816 ACTGCGCAGAGACAGCCCAAACTTGGGCTTTATTTTGGGA 1857

RESULT 13

US-10-120-145-7
 ; Sequence 7, Application US/10120145
 ; Publication No. US20030157684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jefferson, Richard A.
 ; APPLICANT: Kilian, Andrez
 ; APPLICANT: Keese, Paul Konrad
 ; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
 ; TITLE OF INVENTION: USES THEREOF

; FILE REFERENCE: 190106.405
 ; CURRENT APPLICATION NUMBER: US/10/120.145
 ; PRIOR FILING DATE: 2002-04-11
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
 ; NUMBER OF SEQ ID NOS: 71
 ; SOFTWARE: Patent Ver. 2.0
 ; SEQ ID NO: 7
 ; LENGTH: 1887
 ; TYPE: DNA
 ; ORGANISM: Bacillus sp.
 US-10-120-145-7

Query Match 8.4%; Score 141.4; DB 15; Length 1887;
 Best Local Similarity 45.4%; Pred. No. 3.9e-34;
 Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

Qy 120 CCAAGATCTGTGCTACGAAAGAGACCTTACCTACAAACCACTTCTAGCTCGAA 179
 Db 208 CAAGAAATCCGAAACCATATCGATATGTCTGTGCAAGAGTTCACGGTCCGCGC 267
 Qy 180 GNACTTTCAAAAACATCAGACTTATCTTGTGGGTGAGACAGGACTCGAGGT 239
 Db 268 CTATCTGAAGATACAGCTATGTGTCTCCGCTTGGCTTGCATCACAAGCAATTT 327
 Qy 240 CTTCCTCAACGAGAGAAAGTGGAGAAATCATTTGAATACCTTCCCTCGAATAGA 299
 Db 328 CTATCTCAATGTGAGACTGTGTGTGGAGCAAGAGGCGGATTCCTGCTTGAAGCGGA 387
 Qy 300 TGTGAGGGGAAAGTAATCCGGAGAGAACCACTCAGGGTGTGTGGAACAGATT 359
 Db 388 AATCAACAACTGCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 447
 Qy 360 GAAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCAACCGTGGATTTT 419
 Db 448 CGACGATACACCCCTCCGCTGGGTGTACAGCCAGCCACGAAAGGCGCTGGAAA 507
 Qy 420 TGAAGTTTCCACCTGCAAACTTGAATCTTCCCTTACGCTGAAATCAAGAGCTGT 479
 Db 508 AGTATTCGTAAACAGCCCACTTCACTTCAATGATGAGGCTTCAACCTGCTGCT 567
 Qy 480 TCTGATAGATTCAACAGACCAAGAGTACTGCAATCTGGGTGAGACAGAGATGATC 539
 Db 568 GAAATCTACAGACCCGCTTACGATCTGAGAGCAATCTGGGTGAGACAGACTTCA 627
 Qy 540 TGAACCGAAGAACTTTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 599
 Db 628 TGGCCCAACCGGAGCTGTGACTTATACGTGGACTTTCAAGCAGAGAGAGAGCTGTA 687
 Qy 600 GGAACAGAGATGAGATCAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
 Db 688 AGTGTGGGTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
 Qy 660 ATTGTGAG 719
 Db 748 CGTGAAGATTCGAGATGTATCTGTGGAGACCACTGAACAGATCTCTACCAATCA 807
 Qy 720 TCTTATCTCTCAAGGTGAACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
 Db 808 AGTGAATCTGTGAG 867
 Qy 780 AACGATCAGCTGGAG 839
 Db 868 GACGCTGAGATCAAG 927
 Qy 840 CTTTGAAG 899
 Db 928 CTTTGGCAACATGAG 987
 Qy 900 AAAAGACTTCAACTTCTGAAGGTGATCAACGCAATTTCTTTCAGAGACTTCACTATCC 959

```

Db      988 GATGATTTCAATATCTCAAAATGATGGCGCAACAGCTTCGGAACCGACATATCC 1047
Qy      960 TTACAGTGAAGAGTGGCTGATCTTGGCGACAGATCGGAATCTTTGTATGAGAGAC 1019
Db      1048 GTACTCTGAAGATGATGCTCTTGGGATCGCGAGGGTCTGTGATGACGAGAC 1107
Qy      1020 CCCGACGTTGTTAT----- 1034
Db      1108 TCCGGCAGTTGGCGTGCACCTCAATTCATGCGCACACGGGACTCGCGGAAGGACCGA 1167
Qy      1035 ---CAAGAGTACCACTACATCCGAGACTCAGAAATGAGAGAGAACATTAAG 1091
Db      1168 GCGGCTCAGTACTCGGAGAAATTCGAGCTTTGAGACCATCAAGAGCTTCCTCGTA 1227
Qy      1092 AATGATCGACAGACAGAAACCATCCAGTGTATCATGTGAGTGTGGCAAGAAC 1151
Db      1228 ACTGTGTCTGTGACAGAACATCAACGCTCGTGTGATGTGAGATGTGCCAAGAGC 1287
Qy      1152 AGAGTCCAAACATCCAGACGCGGAGGTTCTTCAAGCCCTTATGAGACTGCCATGA 1211
Db      1288 GCGACTGAGAGAGAGGCGGCTGATGATCTTCAAGCCCTTGTGAGCTGACCAAGGA 1347
Qy      1212 AATGATGGA---ACAGCCCGCTTGTCAATGTGAGCATGATGACGACACAGAGAG 1268
Db      1348 ACTGACCCACAGAAAGCGTCCGCTACGATCGTGTGTTGTGATGGCTAACCCGAGAC 1407
Qy      1269 AACAAAGACGTGGCGCTGAAATCTTCAATCGTGTGTGAACAGTACTACGCGCTG 1328
Db      1408 GGACAAAGTCGCCGAACTGA-----TTGACGTCACTCGGCTCAATCGCTATAGCGATG 1461
Qy      1329 GTACATCTTATCAGGGAAGATGAAAG---GACTCAAGCTCGGAAGAAAGACATGA 1385
Db      1462 GTACTTCGATGGCGGTGATCTGAAAGCGGCAAAAGTCATCTCCGCAAGAAATTCACGC 1521
Qy      1386 AGAGCTTATGACAGGACAGAAAGCCCATCTTGTGACAGAAATTCGATGCGGACGCGAT 1445
Db      1522 GTGGAACAAGCGTTCGCGGAAAGCCGATCATGATCACTGATGACGCGGACAGACCGT 1581
Qy      1446 AGCTGGCATCTCAATGATCACTCAATGTTCTTCCGAAAGTACCAAGACAGCTGCT 1505
Db      1582 TCGCGGCTTTCAGCATGATGATCAATGTTTCCAGCGGAATATCAAGTGAATACTA 1641
Qy      1506 TGAAGAAGCATCAGGCTCTTTGAAAAAAGATCAATCATCGGAACAAGTGTGGC 1565
Db      1642 CCAGCGGAACCACTGCTGTGATGATGATGAACTTCGTGGTGAAGCGGTGA 1701
Qy      1566 CTTTCAGATTTTAAGACTCTCAGAAATGTGAGAAAGACCATTCACCAAGAGGTGT 1625
Db      1702 CTTGCGGACCTTCGGAACCTCTCAGGGCGTGTGCGCGCTCAAGGAACAGAAAGGCGT 1761
Qy      1626 TTTCAAGAGACAGACAAACCAAACTGCTGTCTCATGTACTGAGAAAGCTGTGA 1681
Db      1762 GTTCACTGTGACCGCAAGCCGAACTGCGCGCAGACGCTTTCCGAGCGCTGA 1817

```

RESULT 14
US-10-322-656-45

```

; Sequence 45, Application US/10322656
; Publication No. US20030182691A1
; GENERAL INFORMATION:
; APPLICANT: Robert, Laurian
; APPLICANT: Gleddie, Stephen
; APPLICANT: Foster, Elizabeth
; TITLE OF INVENTION: Modification of Pollen Coat Protein Composition
; FILE REFERENCE: 100210.01
; CURRENT APPLICATION NUMBER: US/10/322,656
; PRIORITY FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: 09/272,204
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 60/078,728
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin version 3.0

```

```

; SEQ ID NO 45
; LENGTH: 3651
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: nucleotide sequence of the translational fusion in plasmid AT0G-4
US-10-322-656-45

Query Match      8.4%; Score 141.4; DB 15; Length 3651;
Best Local Similarity 45.4%; Pred. No. 5.7e-34;
Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;

Qy      120 CCAGATCTGTCTACGAAGAAGACCTTCACCTTCAAAACCACTTCTACGTTCCGA 179
Db      1983 CAAGAAATCCGCAACATATCGGATATGTCTGATGACAACTGTATCACTGATCCGCGC 2042
Qy      180 GMAACTTTCACAAAACATCAGACTTACTTCTGTGGGTGAACAGGACTGCGAGGT 239
Db      2043 CTATCTGAAGATCAGCTATCTGCTCGCTTCGCTCTGCACTCACAAAGCAATGT 2102
Qy      240 CTTCTCAACGAGAGAAAGTGGAGAAATCATATTAAATCTTCCCTTGAAGTGA 299
Db      2103 CTATGTCAATGTGTAGTGTGTGTGAGACACAAAGGCGGATTCCTGCACTTGAAGCGGA 2162
Qy      300 TGTGACGGGAAAGTGAATCCGAGAGACGAATCTAGGCTGTGTTGAGAACAGATT 359
Db      2163 AATCAACAACCTGCTGCTGATGAGCATATCGCTCACCGTCCGCTGGAACATCTCT 2222
Qy      360 GAAAGTGGAGAGATTTCCTCGAAGTTCCAGACAGCGGCACTGACACGCTGGATTGTT 419
Db      2223 GAGCATATGACCTCTCCGCTGGGCTGTACAGCCAGCGCCAGAAAGGCGCTCGAAA 2282
Qy      420 TGAAGTTTCCACCTGCAAACTTCGACTTCTTCCCTTACCGTGTGAATCATTAAGCTGT 479
Db      2283 AGTATCTGTAAACAAGCCGAACCTTGCACTTCTTCAATATGAGAGCTGACACCTCGGT 2342
Qy      480 TCTGATTAAGTTCAACAGACACCGGAGATCTGACATCTGGGTGAGACAGATGATG 539
Db      2343 GAAATCTTACACGACCCGTTTACCTACGTGAGAGACATCTCGTGTGAGACGCACTTCA 2402
Qy      540 TGAACCGGAAGAAATCTTGAAGATGAAGTAAATGAAATGATGAAAGAGCCGT 599
Db      2403 TGGCCCAACCGGACTGTGACCTTATACGTGATCTTCAAGGCAAGCCGATGAA 2462
Qy      600 GGGACAGAGATGACATCAAACTTGGAGAGAGAGAAAGAAATTAAGATCCAAAG 659
Db      2463 AGTGTGCTGTGATGAGAGAGAGAAAGTGTGTGCAAGCACCGAGGCGTGAAGGTTAA 2522
Qy      660 ATTGCTGAAGGGAGTTTCATCTCTGAAAACGCCAGTTCTGGAGCTTGAAGATCAATA 719
Db      2523 CGTGAAGATTCGGAATGATCATCTCTGGGAAACCACTGAACAGTATCTTCAACATGA 2582
Qy      720 TCTTATCTCTCAAGGTGGAACCTTGAAGAGAGATGACACTGGAACATCGGAATCAG 779
Db      2583 AGTGAACCTGTGAACGAGACGATGACCATGATGCTTATGAAGAGCCGTTGGGCTGCG 2642
Qy      780 AACGATCAGCTGGAGAGAGAGAGCTTATCTGAACGGGAAACCTGCTTTTGAAGCG 839
Db      2643 GACGCTGAAGTCAACAGACGCGCAAGTCTCTATCAACAAACCGCTTCTACTTCAAGGG 2702
Qy      840 CTTTGAAGAGACAGAGATTCCTCCGCTTGGGAGGAGCACTTTTATCAATGATAT 899
Db      2703 CTTTGGCAAAATGAGGACACTCTCATCAAGCGCGCTGTATTAAGAAAGCAAGATGT 2762
Qy      900 AAAAGACTTCAACCTTGAAGGTGATCAAGCGCAATCTTTCAGGACTCTCAATATCC 959
Db      2763 GATGATTTCAATATCTTCAATGATGAGGCGCAACGCTTTCGAGCGCACTATATCC 2822
Qy      960 TTACAGTGAAGAGTGTGATCTTGGCCGACAGACTCGGAATCTTGTGATAGCAAGAC 1019
Db      2823 GTACTCTGAAGAGTGAATGCGCTTTGGGAGATCGGAGGATCTGATGATGACAGAC 2882

```


Db 376 GCGCGTCAGTACCTGGAGAGATTGCGAGCTTTGAGCAGCATCAAGACGTTCTCGTGA 3435
Qy 1092 AATGATCGACAGACACAGAAACCATTCCTCGTGTGATCATGTGGAGTGGCGAAC 1151
Db 3436 ACTGGTGTCTCGTGAACAGAACCATTCAGACGTCGTGATGTGAGCATCGCCACGAGGC 3495
Qy 1152 AGAGTCGAACCATTCAGACGCGGAGGGTTCTTCAAAAGCCCTTTATGAGACTGCCAATGA 1211
Db 3486 GCGGACTGAGAGAGAGGGGCGCTTCAAGTACTTCAAGCCGTTGTGAGACTGACCAAGGA 3555
Qy 1212 AATGATCGA---ACACGCCCGGTTGTCAATGATGAGCATGATGACGCAACGAGAGAG 1268
Db 3586 ACTCGACCCACAGAAAGCGTCGCGTCAAGTCGTCGTTGTGATGGCTAACCCGAGAGAC 3615
Qy 1269 AACAAAGACGTGGCGCTGAATGACTTGCACATCGTGTGTGAACAGGTACTACGGCTG 1328
Db 3616 GGACAAAGTCGCGCACTGA-----TTGACGTCAATCGCGCTCAATCGCTATACGGAATG 3669
Qy 1329 GTACATCTATCAGGGAAGGATGAAGAG---GACTTCAAGCTCTGMAAAAGACATAGA 1385
Db 3670 GTACTTCGATGGCGGTGATCTGAAAGCGGCCAAAGTCCATCTCCGCGAGGAATTTCAACGC 3729
Qy 1386 AAGGCTCTATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGATGCGAGCGCAT 1445
Db 3730 GTGGAACAAAGCGTGGCCAGAAAGCCGATCATGATCACTGAGTACGGCGCAGACACGCT 3789
Qy 1446 AGCTGGCATCCACTACGATCCACTCAATGTTCTCCGAAGATACCAAGCAGAGCTGCT 1505
Db 3790 TGGGGGCTTTCAAGCATTTGATCCAGTGTGTTCAACGAGAAATATCAAGTCAGTACTTA 3849
Qy 1506 TGAAGAAGCATCAGGCTCCTTTGMAAAAGACTACATCATCGGAACACAGTGGGCG 1565
Db 3850 CCAAGCGAACCCAGTCGTGTTGATGAGTTCGAGAACTTCGTGGTGAACCAAGCGTGA 3909
Qy 1566 CTTTTCAGATTTTAAGACTCTCTAGATGTGAAGAAAGCCCATTTTCAACCAAGGCTGT 1625
Db 3910 CTTGCGGACCTTCGGAACCTCTCAAGGGCGTGAAGCGGCTCAAGGAACAAAGAGGGCGT 3969
Qy 1626 TTTCAAGAGACAGACAAACCAACTGCTGCTCATGTACTGAGAAAGCTGTGA 1681
Db 3970 GTTCACTCGTACCCGCAAGCCGAGCTGCGCGCACGTCCTTTCGAGCGCTGA 4025

Search completed: January 24, 2005, 07:49:18
Job time : 970 secs

!!is Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2005, 03:17:30 ; Search time 5564 Seconds
(without alignments)

11061.594 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689
Sequence: 1 atggttaagaccgcgaacga.....gaagactgtgagcagaggtc 1689

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 32822875 seqs, 1821985908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: gb_ear1:
2: gb_ear2:
3: gb_hnc:
4: gb_ear3:
5: gb_ear4:
6: gb_ear5:
7: gb_ear6:
8: gb_ga81:
9: gb_ga82:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174	10.3	1377	3	CR593823 full-length
2	174	10.3	1853	6	CD014093 90135266
3	174	10.3	1995	6	CD014092 90134967
4	161.6	9.6	1124	6	CD503076 CDA60-C07
5	161.2	9.5	1138	6	CD503098 CDA60-D07
6	159.6	9.4	2274	3	AK041058 Mus muscu
7	150.6	8.9	816	7	CK467358 938649 MA
8	144.4	8.5	906	5	BQ941196 AGENCOURT
9	142.6	8.4	878	7	CN155220 942501 MA
10	141.6	8.4	1051	5	BX363460 BX363460
11	133.6	7.9	767	6	CF255373 mdvnl27 c
12	131	7.8	846	6	CA453907 AGENCOURT
13	128.8	7.7	900	6	CA489317 AGENCOURT
14	128.8	7.7	914	6	CB203472 AGENCOURT
15	128.2	7.6	862	7	CO394571 AGENCOURT
16	126.4	7.5	2473	3	AY321342 Rattus no
17	121	7.2	857	5	BX745933 BX745933
18	117.4	7.0	1287	6	CD503099 CDA60-D07
19	117.2	6.9	802	5	BUS57212 603474094
20	116	6.9	725	5	BUS15788 603851444
21	114.2	6.8	724	5	BM441165 BM441165
22	114.6	6.7	771	4	BG121498 602352830
23	111.6	6.6	790	7	CO559438 AGENCOURT
24	110.8	6.6	740	7	CF521612 AGENCOURT

25	109.6	6.5	752	4	BJ733141
26	109.4	6.5	1268	6	CD503077 CDA60-C07
27	107.6	6.4	673	6	CB841568 M15E-2121
28	107.4	6.4	791	5	BUS239978 603323761
29	107.2	6.3	915	5	BQ678153 AGENCOURT
30	106.6	6.3	543	5	BQ360163 QVO-OT003
31	106.6	6.3	615	4	BG568800 602588157
32	106.4	6.3	975	4	BG285429 602406582
33	105.6	6.3	616	2	BE179118 RC0-HT061
34	105.4	6.2	504	5	BUS17201 EST143 Bo
35	105.2	6.2	736	5	BP454056 BP454056
36	104.6	6.2	757	4	BI851705 603377940
37	104	6.2	877	7	CF406520 CH3042 G
38	101.8	6.0	1023	5	BX363459 BX363459
39	101.6	6.0	773	7	CF748482 UI-M-HUO
40	101	6.0	777	4	BG741172 602631852
41	100.8	6.0	667	7	CN264045 CN264045
42	100.6	6.0	548	2	BE751040 202745 MA
43	100.2	5.9	799	5	BUS66987 603585124
44	100	5.9	701	5	BX611242 BX611242
45	99.6	5.9	921	9	CL486845 SATL_443_

ALIGNMENTS

RESULT 1
CR593823
LOCUS full-length cDNA clone CS0DL001W21 of B cells (Ramos cell line)
DEFINITION Cor 25-normalized of Homo sapiens (human).
ACCESSION CR593823.1 GI:50474630
VERSION CR593823.1
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1377)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@life.techn.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
2 (bases 1 to 1377)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr
COMMENT 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
source location/Qualifiers
1. 1377
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL001W21"
/tissue_type="B cells (Ramos cell line) Cor 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 10.3% Score 174; DB 3; Length 1377;
Best Local Similarity 52.2% Pred. NO. 7.3e-40;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
Oy 737 TGGAACTTGAAGAAAGACAGTACACTGTGACATCGATGATGAGACGATGAGTGGAGCG 796
Db 215 TGGGGCCGTGTGCTGACCTTCTACACACTCCCTGTGGGATCCGACCTGTGGCTGCACCA 274

```

QY 797 AGAAGAGCTTATCTGAAACGGGAAACCTGCTTTTGAAGGCTTTGAAAACGCGAG 856
Db 275 AAGAGCAGTCTCTCAATCAATGGGAAACCTTTCTATTTCCACGGTGTCAACAGCATGAG 334
QY 857 AATTCCTGCTTGGGGGAGGCGACCTTTATCATTTGATGATAAAGACTTCAACCTTC 916
Db 335 ATGCGGACATCCGAGGAAAGGGCTTCACTGCGCCCTCTGTGTGAAGACTTCAACCTGC 394
QY 917 TGAAGTGAATCAAGCGGAATTTTTCAGACCTTCACTATCTTCAAGTGAAGTGGC 976
Db 395 TTCGTGGCTGGTGTGCAACGCTTTCGATCCAGGCACTACCTTATCAGAGGAGTGA 454
QY 977 TGGATCTTGCCGAGACGACTCGGAATCTTGTGATGAGAAAGCCCGGACGCTGTGATCA 1036
Db 455 TGCAATGTGTGACCGCTATGGGATTTGTGTATGATGATGATGATGATGATGATGATGAT 514
QY 1037 CAAGGTACCA---CTACAAATCCGAGACTCAGAAAGATAGCAGAAAGACATTAAGAA 1093
Db 515 CGCTCCCGCAGTTCTTCAACAAACGTTTCTTGCATCACCACATGCAAGGTATGAGAAAG 574
QY 1094 TGAATGACAGACAAAGAACCATCCCACTGATGATGATGATGATGATGATGATGATGATGAT 1153
Db 575 TGGTGGCTGAGGACAAAGAACCAACCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 634
QY 1154 AGTCCAACTATCCAGACGCGAGGGTCTTCAAAAGCCCTTATGAGACTGCGCAATGAAA 1213
Db 635 CGTCCACACTTGAATATCTGCTGCTATCTTGAAGATGTATGCTTCAACCAATCTT 694
QY 1214 TGAATGACAGACGCGCGCTGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1273
Db 695 TGGACCCCTCCCGGCTGTGACCTTTGTGACCACTTATGATGATGATGATGATGATGATGAT 742
QY 1274 GAGAGCTGGCGCTGAGTACTTGCACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1333
Db 743 CAGACAAAGGGGCTCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 802
QY 1334 TCTATCAGGGAAGATAGAAAGAGACTTCAAGCTTGTGAAAAGACATTAAGAAAGCTT 1393
Db 803 ACGACTAGGGGCACTGAGTTGATTCAGCTGCACTGCGCAACCACTTGTGAAGACTGT 862
QY 1394 ATGCAAGGACAAAGAGCCCATCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1453
Db 863 ATAAAGATATCAGAAAGCCATTTATCAGAGGATGTGAGAGCAAGAAACGATTTGAGGT 922
QY 1454 TCCACTAGGATCCACTCAATGTTCTCCGAAAGTACCAAGGAGACTGTTGAAGAG- 1512
Db 923 TTCAACGAGATCCACTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 982
QY 1513 ACGATCAGGCTCTTTGAAAAA---AGACTATCATCTGAAACACAGTGTGGGCT 1567
Db 983 ACGATCTGGGCTGGATCAAAAAACGAGAAATACGTTGAGTGTGAGTCACTTTTGAAT 1042
QY 1568 TTGAGATTTTAAATCTCTGAAATGTGAGAAAGACCATTTTCAACCAAGGCTTTT 1627
Db 1043 TTGCGGATTTTATATCTGAAGAGTACCGAGAGAGTGTGGGAAATTAAGAGGAGATCT 1102
QY 1628 TCACAAAGACAGACCAACCACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1671
Db 1103 TCATCGGAGAGACCAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1146

```

RESULT 2
 LOCUS CD014093 1853 bp mRNA linear EST 21-OCT-2003
 DEFINITION 90155266 single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014093
 VERSION CD014093.1 GI:37777622
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE 1 (bases 1 to 1853)
AUTHORS Jin, P., Fu, G.X., Wilson, A.D., Yang, J., Chien, D., Hawkins, P.R.,
Au-Young, J., and Stuve, L.L.
TITLE PCR isolation and cloning of novel splice variant mRNAs from known
JOURNAL drug target genes
COMMENT Genomics 83 (4), 566-571 (2004)
CONTACT: Jin, P.
Incyte Corporation
3160 Porter Drive, Palo Alto, CA 94304, USA
Tel: 650 621 8639
Fax: 650 621 8965
Email: pj@incyte.com.
FEATURES
source
1..1853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="single gene library"
/note="Vector: pDrive Cloning Vector; RT-PCR was performed
using gene-specific primers flanking the open-reading
frame. PCR products were subcloned into pDrive Cloning
vector and sequenced completely using M13 forward and
reverse primers. Sequencing gaps were closed by
re-sequencing using primers flanking the gapped areas."
ORIGIN
Query Match 10.3%; Score 174; DB 6; Length 1853;
Best Local Similarity 52.2%; Pred. No. 8.2e-40;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
QY 737 TGAAGCTGAAAAAGACGAGTACACTCTGAGATCATGAGATCGAAGATCTGAGAG 796
Db 624 TGGAGCTGTGTGTGCTTCTTCAACACTCCCTGTGGGATCGCACTGTGCTGCACA 683
QY 797 AGAAGAGCTTATCTGAAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGACAGAG 856
Db 684 AGAGCAAGTTCCATCATATGAGAAACCTTTCTATTTCCAGGTGTCAACAGATGAG 743
QY 857 AATTCCTGCTTGGGGGAGGCACTTTTATTCATTTGATGATTAAGAACTTCAACTTC 916
Db 744 ATGCGGACATCCGAGGAAAGGGCTTCACTGCGCTGTGTGTGAAGACTTCAACTTC 803
QY 917 TGAAGTGAATCAACGGGAATCTTTCAGGACCTCTCACTATCTTACAGTGAAGTGGC 976
Db 804 TTGCTGCTGTGTGTGCAACGCTTTCCTTCAACGCACTACCTTATGCAAGAGAGTGA 863
QY 977 TGAATCTCCGACAGACTCGGAATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1036
Db 864 TGCAGATGTGTACCGCTATGGGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 923
QY 1037 CAAGGTACCA---CTACAAATCCGAGACTCAGAAAGATAGCAGAAAGACATTAAGAA 1093
Db 924 CGCTGCGGAGTTCTTCAACAAAGTTTCTTGCATCACCAATGCAAGGTATGAGAAAG 983
QY 1094 TGAATGACAGACAAAGAACATCCCACTGATGATGATGATGATGATGATGATGATGATGATGAT 1153
Db 984 TGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1043
QY 1154 AGTCCAACTATCCAGACGCGAGGGTCTTCAAAAGCCCTTATGAGACTGCAATGAAA 1213
Db 1044 CGTCCACCTAGAAATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1103
QY 1214 TGAATGAAACAGCGCCGTTGTGATGTGAGATGATGAGAGCAACCAAGAGAGAA 1273
Db 1104 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCACTTAACTATGACAG----- 1151
QY 1274 GAGAGTGGCGCTGAGTACTTGCACATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1333
Db 1152 CAGACAAAGGGGCTCCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1211
QY 1334 TCTATCAGGGAAGATTAAGAGAGACTTCAAGCTTGTGAAAAGACATTAAGAGCTCT 1393
Db 1212 ACGACTAGGGGCACTGTGATGATTCAGCTGCACTGGCAACCAAGTTTGAAGAACTGT 1271

```


QY 1394 ATGCAAGGACAGAAAGCCCATCTTGTTCACAGAAATTGCGTCCGACCGGATAGCTGCA 1453
 DB 1272 ATTAAGATGATCAGAAAGCCCATCTTGTTCAGAGCGAGATGAGACAGAAAGATTTGACGGGT 1331
 QY 1454 TCCACTAGATCACCCTCAATATGTTCTCCGAAAGATACCAAGACAGAGCTGTTGAAAAG- 1512
 DB 1332 TTCACAGAGATCCACTCTGATGTTTCACTGAAAGATACCAAGAAAGTCTGCTAGAGCACT 1391
 QY 1513 ACGATCAGGCTCTTTTGA AAAA-----AGACTACATCATCGAAACACAGCTGTGGGCT 1567
 DB 1392 ACCACTGTGGTCTGATCAAAAACGAGAAATACGTGGTTGAGAGCTCATTTGAAATT 1451
 QY 1568 TTGCAGATTTTAAAGCTCCTCGAATGTGAGAAAGCCATTTCTCAACCAAGGCTGTT 1627
 DB 1452 TTGCCGATTTCTGACTGAAAGTCAACGACGAGAGTGTGGGAAATMAAAAGGGAGTCT 1511
 QY 1628 TCACAAAGACAGAACCCAAACTGTTGCTCATGTACTGAGA 1671
 DB 1512 TCACCTGGCAGAGACAAACCAAAAGTGCAGCTTCTTTGGCA 1555

RESULT 3
 LOCUS CD014092 1995 bp mRNA linear EST 21-OCT-2003
 DEFINITION 90134967 Single gene library Homo sapiens cDNA, mRNA sequence.
 ACCESSION CD014092
 VERSION CD014092.1 GI:37777621
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1995)
 JIN, P., FU, G. K., WILSON, A. D., YANG, J., CHEN, D., HAWKINS, P. R., AU-YOUNG, J. and STUVE, L. L.
 PCR isolation and cloning of novel splice variant mRNAs from known drug target genes

TITLE Genomics 83 (4), 566-571 (2004)
 JOURNAL Contact: Jin, P.
 COMMENT Incyte Corporation
 3160 Porter Drive, Palo Alto, CA 94304, USA
 Tel: 650 621 8639
 Fax: 650 621 8965
 Email: pjin@incyte.com.
 Location/Qualifiers

FEATURES
 Source 1..1995
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed using gene-specific primers flanking the open-reading frame. PCR products were subcloned into pDrive Cloning Vector and sequenced completely using M13 forward and reverse primers. Sequencing gaps were closed by re-sequencing using primers flanking the gapped areas."

ORIGIN

Query Match 10.3%; Score 174; DB 6; Length 1995;
 Best Local Similarity 52.2%; Pred. No. 8.4e-40;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAAGACGATCACTGTGACATCGAATCGAAGCATCACTGAGAGC 796
 DB 766 TGGGGCTGTGTCTGATCTTACACATCCCTGTGGGATTCGCACTGTGCTGCACCA 825
 QY 797 AGAAGAGCTATCTGAACGGAAACCTGCTTTTGAAGGGCTTTGAAAGCAGAG 856
 DB 826 AGAGCAGTCTCTCAATGAGAAACCTTTATTTCCACGCTGTCAACAGCATGAG 885
 QY 857 AATTCCTGCTTGGGGAGGACACCTTTATTCATTTGATGATTAAGACTCACTTC 916

DB 886 ATGCGACATCCAGAGGAGGCTTGCATGCGCGCTGCTGGTGAAGACTTCAACTGC 945
 QY 917 TGAAGTGAATCAAGCGCAATTTCTTACAGACTCTCATATCTTACAGTGAAGTGC 976
 DB 946 TTGCTGAGCTTGTGCTCAACGCTTCCGTACAGACCATACCCCTATCAGAGGAATGA 1005
 QY 977 TGGATCTTCCGACAGACTCCGGAATCTTGTGATGAGAAAGCCCGACCTGTGATCA 1036
 DB 1006 TGCAGATGTGACCGCTATGGAATGTGTGATCATGATGATGATCCCGCGGCGCTGG 1065
 QY 1037 CAAGTACCA---CTACATCCGAGACTCAGAAATGACAGAAACAAATTAAGAA 1093
 DB 1066 CGCTCCGCAATTTCTTCAACAAGTCTCTGATCAACACATGAGAGTGAAGAAAG 1125
 QY 1094 TGATGACAGACACAAAGCAATCCCACTGATCATGTGATGATGATGATGATGATGAT 1153
 DB 1126 TGTGCTGTGAGGACAAAGAACACCCCGGCTGTATGTGTGTGTGTGTGTGTGTGTGT 1185
 QY 1154 AGTCAACCATCCAGCGGAGGCTTCTTCAAGCCCTTATGAGCTGCAATGAAA 1213
 DB 1186 GTTCCACTTGAATCTGCTGCTACTACTTGAAGATGATGATGATGATGATGATGATGAT 1245
 QY 1214 TGGATCGAACAAGCCCGCTTGTCTATGATGATGATGATGATGATGATGATGATGAT 1273
 DB 1246 TGGACCCCTCCCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1293
 QY 1274 GAGAGTGGCGCTGAAGTACTTGCACATGCTGTGTGTAACAGTACTTACGCTGTGTA 1333
 DB 1294 CAGACAAAGGGGCTCCGTATGTGATGATGATGATGATGATGATGATGATGATGATGAT 1353
 QY 1334 TCTATCAGGAAAGATTAAGAAAGATTTCAAGCTCTGGAAGAAACATTAAGAGCTCT 1393
 DB 1354 ACGACTAGCGGACCTGGAAGTGTATGATGATGATGATGATGATGATGATGATGATGAT 1413
 QY 1394 ATGCAAGGACAGAAAGCCATCTTGTTCACAGAAATTGCGTCCGACCGATAGCTGCA 1453
 DB 1414 ATTAAGATTAACAGAAAGCCATTTATTCAGAGGATATGAGACAGAAAGATTTGAGG 1473
 QY 1454 TCCACTAGATTCACCTCAATATGTTCTCCGAAGTACCAAGACAGAGCTCGTTGAAAG- 1512
 DB 1474 TTCACAGAGATCCACTCTGATGTTTCACTGAAAGTACCAAGAAAGTCTGTAGAGCAGT 1533
 QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATGGAACACAGCTGTGGCCT 1567
 DB 1534 ACCACTGTGGTCTGATCAAAAACGAGAAATAGTGTGGAAGCTCATTTGGAAATT 1593
 QY 1568 TTGCAGATTTTAAAGCTCCTCAAGATGTGAGAAGCCATTTCTCAACCAAGAGGTGTT 1627
 DB 1594 TTGCCGATTTTATGATCACTGAACGATCAACGAGAGAGTCTGGGAAATMAAAAGGGATCT 1653
 QY 1628 TCACAAAGACAGAACCCAAACTGTTGCTCATGTACTGAGA 1671
 DB 1654 TCACCTGGCAGAGACAAACCAAAAGTGCAGGCTTCTTTGGCA 1697

RESULT 4

LOCUS CD503076 1124 bp mRNA linear EST 12-JUN-2003
 DEFINITION CDA60-C07.x1d-t SHG-CDA Gasterosteus aculeatus cDNA clone
 CD503076
 ACCESSION CD503076
 VERSION CD503076.1 GI:31432977
 KEYWORDS EST.

SOURCE Gasterosteus aculeatus (three spined stickleback)
 ORGANISM Gasterosteus aculeatus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes; Gasterosteidae; Gasterosteus.
 1 (bases 1 to 1124)
 KINGSLEY, D. M., PEICHEL, C., BALABHADRA, S., GRIMWOOD, J., DICKSON, M., SCHMUTZ, J. and MYER, R. M.

TITLE Expressed sequence tags from Gasterosteus aculeatus

JOURNAL
COMMENT

Unpublished (2003)
Contact: Kingley, DM
HHMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingley@cmgm.stanford.edu
Plate: 60
High quality sequence stop: 782.
Location/Qualifiers
1..1124

FEATURES
source

/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA60-C07"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with a 50 bp linker primer containing an oligo dt sequence preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector undirectionally in the sense orientation with respect to the lacZ promoter of pBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual pBK-CMV phagemid clones for EST sequencing."

ORIGIN

Query Match 9.6%; Score 161.6; DB 6; Length 1124;
Best Local Similarity 51.5%; Pred. No. 3,46-36;
Matches 461; Conservative 0; Mismatches 414; Indels 21; Gaps 3;

QY 748 AAGAGCGTCACTCTTGAAGAGGCTTTGGAAGACGAGAACTCCCGT 867
DB 71 AAGACGCTACCTTACAGTCCGACGCGTTCACGACGCTTACGACGACCACTTC 130
QY 808 TATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGACGAGAACTCCCGT 867
DB 131 CTATCAACAAAGCCCTTCTACTTCCAGAGTAATTAACGAGACTTGATATT 190
QY 868 CTGGGGCAGGGCACCCTTTATCATGATGATTAAGACTTCACTTGAAGTGATC 927
DB 191 CGAGGCAAGGCGCTGAGCTGCGCCCTCATGGTGAAGACTTAACTTATGAAGTGTG 250
QY 928 AAGCGAATTTCTCAGAACTCTCATCTTCACTTCAAGTGAAGGTGGCTTGGC 987
DB 251 GGGGCACTCTGTCGACACGACCTTATGCAAGAGGATCTCGAGATGTGT 310
QY 988 GACAGACTCGGAATCTTGTGATAGAGAACGCCGACCTGTGTATCACA--AGGTAC 1044
DB 311 GACCGCATGGCATGTGTGATAGAGAGTCCCGGCGGTGAGCAATTCGCG 370
QY 1045 CACTTCAATCCGAGACTCAGAGATAGCAGAAACAATTAAGAAATGATGACAGA 1104
DB 371 AGTTTGAAGACGCTCTTAACCATCACCTGTCTCATGACGAGCTGTAAGTGG 430
QY 1105 CACAAGAACATCCAGTGTGATCATGTGAGTGTGCGAAGACGAGTCCAAACAT 1164
DB 431 GACAGAAACATCCCTCTGTGTGATGTGCTGAGTGCAGATAGCGCGTCAAGATG 490
QY 1165 CCAAGCGCGAGGGTTTCTTCAAGCCCTTATGAGATCCCAATGAATGATGACACA 1224
DB 491 CTCCTCTGATTAATTTCAAAACCTTGATTAAGATCAACAAAGATTTGATCCAAAC 550

QY 1225 CGCCCCCTTGTCACTGTGACATGATGAGACGACAGACGAGAAACAGAGACGTGGC 1284
DB 551 CGGCCCT-----CACTTATCAGACAGATTAATATGACAGGATTAAGGG 598
QY 1285 CTGAAGACTTTCAGACATCGTCTGTGAAAGAGTACTACGGCTGTTCATCTATCAGGA 1344
DB 599 GCTCCCTAGGTGACATCATCTCGTAAAGTAACTTCTCTGTGATCCATGACCCGGGC 658
QY 1345 AGATAGAAAGAGACTTCAAGCTCTGAAAGAAAGATAGAAAGCTTATGCAAGGAC 1404
DB 659 CACCGGAGGTCAATCCCATCCAGCTCAACATCATGTTGAGAACTGTACGAAAGTAC 718
QY 1405 AGAAGCCATCTTTGTCAAGAAATTCGTGCGGACGCCATAGCTGATCCATACAT 1464
DB 719 CAGAAACCATCATCCAGACGGAATACGAGCGGATGCGGTGCGGGGCTTCAAGTAT 778
QY 1465 CCACTCAATGTCTTCCGAAAGTACCA-----AGCAGAGCTGTGAAAGAGATC 1518
DB 779 CACCCGTGATGTTTACTGAGAGGTACCAAGATTGCTCTGACAGCTACCAACGTG 838
QY 1519 AGGCTCTTTTGAAGAAAGCTATCATTCGAAACACAGTGTGGCTTTGCAATTT 1578
DB 839 TTGACACAGAAAGAGAGCATGTCATCGGGAATCATCTGAACTTTGACAGCTTC 898
QY 1579 AAGACTCTCAGATGTGAAGAACCCATTTCAACCAAGGTTTTCACAG 1634
DB 899 ATGACCCACAGAGGATCATGCGGTGTGGGAGCAAAAGGTTCTCTCAGACAG 954

RESULT 5
CD503098 1138 bp mRNA linear EST 12-JUN-2003
DEFINITION CDA60-D07.x1d-t SHGC-CDA Gasterosteus aculeatus cDNA clone
LOCUS CD503098
ACCESSION CD503098
VERSION CD503098.1 GI:31433163
KEYWORDS EST.
SOURCE Gasterosteus aculeatus (three spined stickleback)
ORGANISM Gasterosteus aculeatus
Bukayoka; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.
1 (bases 1 to 1138)

REFERENCE
AUTHORS Kingley, D.M., Petchel, C., Balabhadra, S., Grimwood, J., Dickson, M.,
Schmutz, J., and Myers, R.M.
Expressed sequence tags from Gasterosteus aculeatus
Unpublished (2003)
CONTACT: Kingley, DM
HHMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739
Email: kingley@cmgm.stanford.edu
Plate: 60

High quality sequence stop: 782.
Location/Qualifiers
1..1138

/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA60-D07"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/note="Vector: lambda ZAP Express/pBK-CMV; Site 1: EcoRI (5' adaptor); Site 2: XhoI (3' linker primer); The mixed organ cDNA library was generated using the ZAP-cDNA method by Stratagene. First strand cDNA synthesis was primed with a 50 bp linker primer containing an oligo dt sequence

COMMENT

Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.
URL: <http://genome.gsc.riken.jp/>
URL: <http://fantom.gsc.riken.jp/>
Location/Qualifiers

FEATURES

Source

1..2274
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A530072005"
/db_xref="taxon:10090"
/clone="A530072005"
/sex="male"
/tissue_type="aorta and vein"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
14..1778
/note="beta-glucuronidase structural (MGD|MG1:95874, GB|NM_010368, evidence: BLASTN, 99%, match=733)
putative"
polyA_signal
2260..2265
/note="putative"
2274
polyA_site
2274
/note="putative"

ORIGIN

Query Match 9.4%; Score 159.6; DB 3; Length 2274;
Best Local Similarity 50.6%; Pred. No. 1.7e-35;
Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;

739 GAACCTGAAAGAGAGATACATCTGACATCGGAATCAAGACATGCTGGAGAG 798
759 GAGTCTGACTGACTACTACACCTTCCTTCCGGATTCAGACAGCTGCTGACAAAG 818
799 AAGAGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGCTTTGMAAGACAGAGAA 858
819 AGCAAGTCTCTATTAACGGGAAACCTTCTTATTTCCAGGGGCTCATTAAGCAGAGAT 878
859 TTCCCGCTTCTGGGCGAGGCACTTTATTCATTGATGATTAAGAACTTCAACTTCTG 918
879 TCAGATATCCAGAGGAAAGCTTCGACTGCGCTGCTGTAAGATTTCAACTCTCTC 938
919 AAGTGATCAACGGGAATCTTTTCAGGACCTCTCATCTTCACTGAAAGTGGCTG 978
939 CGTTGGCTCGGGGCAAAATTCCTTCTGTCACGCCATATCTTCAAGAGAGTACTT 998
979 GATTTGGCCGACAGACTCGGAATCTTGTGATAGACCAAGCCCGCAGCTTGG---TATC 1035
999 CAGCTGTGTGACGATACGGGATTTGTGTATGATAGTGTCCGGTGTGGGATTTGTG 1058
1036 ACAAGTATCCACTATCCCGAGACTCAGAAATAGCAGAGCAACATTAAGAGATG 1095
1059 CTACCTCAGAGTTTGGCAAGATCACTTCGGCACCACTAGAGGTGATGAGAGAGCTG 1118
1096 ATCGACAGACAGACAGACATCCCAAGTGTATCATGTGAGTGGGAGAGCAACAGAG 1155
1119 GTTCCGCGGAGCAAAATACACCTCGCGTGTATGTGTCTGTGGCAATGAGCTTCC 1178
1156 TCCACCATCCAGACGGGAGGGTTTCTTCAAGCCCTTTATAGAGATGCAATGAATG 1215
1179 TGTGCTCTGAAACCGCGCGCATATTACTTTAAGCGGTGATCAACCAACCAAGCCCTG 1238
1216 GATGAAACAGCCCGCTTGTATGTGATGATGATGACGACCAACAGAGAGAAACAGA 1275

1239 GACCTCACCCTGCTCCGAGACTTTGTGAGCA-----ACGCCAATATGATGCA 1286
1276 GAGCTGCGCTGAAGTACTCTTGCATCTGCTGTGTAAGAGATCAACGCTGTATCATC 1335
1287 GACTTGGGGGCCCCGTCAGTGTATGTATGTGTAAACGCTACTTTCTTGTATCAT 1346
1336 TATCAGGAAAGATAGAAAGAGACTTCAAGCTCTGAAAAAAGACATAGAAAGCTTAT 1395
1347 GACTATGGCATTTGAGAGGATTCAGACCACTGAATAGCCAGTTTGAACTGGTAT 1406
1396 GAAAGACAGAAAGCCCATCTTGTGCACAAATTTGGTGGGAGCCGATAGCTGCATC 1455
1407 AAGAGCATAGAAAGCCGATTTATTCAGAGAGATATAGAGCAACCAATCCAGAGATC 1466
1456 CACTACATCCACCTCAATGTCTCCGAAGATCAACAGAGCTCGTTGAAAAAGAG 1515
1467 CAGAGAGACCCGCTCGCATGTTCACTGAGAGATACCAAGGCTGTTGAGAAATTAC 1526
1516 ATCAGGCTCTTTTGA---AAAGCTACATCATGGAACACACGTTGGGCTTT 1569
1527 CATTCACTCTGATCAGAAACGTAAGATACGTGTGAGAGCTCATCTGGAATTTT 1586
1570 GCAGTTTAAAGCTCCTCAGATGAGAGAGCCATTTCAACCAAGGCTTTTTC 1629
1587 GCCGACTTCTATGCGAACCGATCACCGCTGAGAGTATGGAACCAAGAGGAGATCTTC 1646
1630 ACAAGACAGACACCAACCACTGTTGCTCATGTACTGAGA 1671
1647 ACTGCCAGAGACGCCAAACTTGCGCTTATTTTTCGA 1688

RESULT 7

CK467358 816 bp mRNA linear EST 14-JAN-2004
938649 MARC 4Pig Sus scrofa cDNA 5', mRNA sequence.

ORIGIN

REFERENCE

Smith, T.P.L., Freking, B.A., Ford, J.J., Vallier, J.L., Wise, T.A., Nissemann, D.J., Wray, J.E. and Keeler, J.W.
Porcine EST collection using a normalized library constructed from embryos representing early developmental stages
Unpublished (2003)
Contact: Smith TPL
USDA, ARS US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

TITLE

Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alc option. Vector identified with cross_match v0.990329.

JOURNAL

place: TMM8042 Row: M Column: 18
Seq primer: GTAATACGACTCATATAGG.
Location/Qualifiers

FEATURES

1..816
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9923"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4Pig"
/note="Vector: pcDNA3.1; site 1: EcoRI; site 2: NotI; library made with combined RNA from day-10, day-13, day-15, day-25, and day-30 whole embryos."

ORIGIN

COMMENT	FEATURES
<p>Contact: Robert Strausberg, Ph.D. Email: cgapbs-rc@mail.nih.gov Tissue Procurement: DCTD/DTF/Gazdar cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L10M2594 row: 9 column: 16 High quality sequence stop: 765.</p>	<p>Location/Qualifiers 1. .906 /oranism="Homo sapiens" /mol type="mRNA" /db xref="taxon:9606" /clone="IMAGE:6420519" /tissue type="large cell carcinoma" /lab host="DH10B (phage-resistant)" /clone_id="NIH_MGC_18" /name="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT. (Life Technologies). Note: this is a NIH_MGC library."</p>
ORIGIN	
<p>Query Match 8.5%; Score 144.4; DB 5; Length 906; Best Local Similarity 52.1%; Pred. No. 4,36-31; Matches 454; Conservative 0; Mismatches 396; Indels 22; Gaps 5;</p>	
<p>828 CTTTGTGAAGGGCTTTGGAAGACGCGGAATTCCTGGGCGAGGGCACCTTTTA 887 1 CTATTTCCACGGTGTCAACAAGCAATGAGATGCGGACATCCGAGGGAAGGGCTTTCGACTG 60</p> <p>888 TTCCATTTATGATTAAGACTTCAACCTTTCGAAAGTGATCAAGCGCAATCTTTCAGGAC 947 61 GCCCGTGTGGTGAAGGACTTCAACCTTTCGCTTGGCTTGGTGCACACCTTTCGGTAC 120</p> <p>948 CTCTCACTATCTTACAGTGAAGATGGCTGATCTTTCGCGACAGACTCCGAATCCTTGT 1007 121 CAGCCACTACCCCTATGACAGAGAGAGATGTGCAATGTGACCCGTATGGAGTTGTGCT 180</p> <p>1008 GATGACGAGAGCCCGGACGTTGGTATCACAGGTACCA--CTACAAATCCGAGACTCA 1064 181 CATGATGATAGTGTCCCGCGTGGCGCTGCGGCTGCCAGATTTCTTCAACAACGTTTCTCT 240</p> <p>1065 GAAGATGACGAGAACAACTAAGAAATGATTCGACAGACAAAGAAACCATCCCACTGT 1124 241 GCATTCACCACTGACAGGTGATGAGAAAGTGTGCGTATGAGGACAAAGAACACCCCGCGT 300</p> <p>1125 GATCATGTGAGTGTGGCGAAGCAAGACAGAGTCCAACTTCACAGACGCGAGGGTTCTT 1184 301 CGTATGTGTGTGTGGCCAAAGAGCTGGCTCCCACTTGAATCTGTGTGCTACTACTT 360</p> <p>1185 CAAAGCCCTTTATGAGACTGCCAATGAAATGATTCGAACACGCCCGTGTTCATGTGTAG 1244 361 GAATGATGTATGCTCTCACACCAATCTTGTGACCCCTCCCGGCTGTGACTTTGTGAG 420</p> <p>1245 CATATGACGCGACCAAGACGAGAGAACAAAGACGTGTGGCGCTGAATGATCTTTCGACATCGT 1304 421 CAACCTTAACATGACG-----CAGACAAAGGGGCTCCGATGTGATGTGAT 468</p> <p>1305 CTGTGTGAACAGGACTACGGCTGTGATCATCTTATCAGAGGAGAGATAGAAAGAACTTCA 1364 469 CTGTGTGAACGCTACTACTCTTGTGTATCAGACTAGCGGCACTGAGATGTGATCACT 528</p> <p>1365 AGCTTGTGAATAAGACTCTATGCAAGGCAAGAAAGCCCATCTTTGTCTAC 1424</p> <p>529 GCACCTGGCCCAACGATTTGGAAGCTGTATTAAGAGATATCAAGAGCCCATTTATCAGAG 588</p>	

QY 1425 AGAATTGGTGGGACGCGATAGTGGATCCACTAGACCTCAATGTTCTCCGA 1484
 DB 589 CGAGTATGAGACAGAAAGATGTCAGGGTTTCACAGATCACCTCGATGTTCACTGA 648
 QY 1485 AGAGTACCAACAGAGCTGTTGAAAAG-ACGATCAGGCTCTTTGAAAAAG-----A 1538
 DB 649 AGAGTACCAAGAAAGTGTGTAGAGCAGTACCATCTGGGTCTGATCAAAAACAGAAA 708
 QY 1539 CTACATCATCGGAACACAGCTGTGGGCTTTGCAGATTTTAAGCTCTCTCAAAATGTGAG 1598
 DB 709 ATAGGTGTTGAGAGCTCATTTTGAATTTTGCCGATTTCTAGCTGAACAGTCACCGAC 768
 QY 1599 AAGA-CCCATCTCAACCAAGAGGTGTTTTCACAAGAGACAGAACCCAACTCGTTG 1657
 DB 769 GAGAGTGTGGGGAAATAAAGGAGATCTTCACTCGGAGACACCAAAAGTGCAG 828
 QY 1658 CTCACTGTAAGAGAGCTGTGAGTGAAGTT 1689
 DB 829 GCTTCCTTTTGCAGAGAGATACCTGGAGATT 860

RESULT 9
 CN155220 878 bp mRNA linear EST 02-Apr-2004
 LOCUS 942501 MARC 4P1G Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION
 ACCESSION CN155220
 VERSION CN155220.1 GI:46169650
 KEYWORDS EST.

SOURCE
 ORGANISM
 Sus scrofa (pig)
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 878)

REFERENCE
 AUTHORS Smith,T.P., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
 Nomenan,D.J., Wray,J.E. and Keele,J.W.
 Porcine EST collection using a normalized library constructed from
 embryos representing early developmental stages
 Unpublished (2003)

JOURNAL
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov
 Single pass sequencing; Bases called with phred v0.020425.c and
 trimmed with the aid of the trim_alt option. Vector identified with
 cross_match v0.990329.
 Plate: TWM8046 row: 1 column: 19
 Seq primer: GTAATACGACTCACTATAGG.
 Location/Qualifiers

FEATURES

source
 1..878
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 4P1G"
 /note="Vector: pCDNA3.1; Site_1: EcoRI, Site_2: NotI;
 library made with combined RNA from day-10, day-13,
 day-15, day-25, and day-30 whole embryos."

ORIGIN

Query Match 8.4%; Score 142.6; DB 7; Length 878;
 Best Local Similarity 52.1%; Pred. No. 1.5e-30;
 Matches 425; Conservative 0; Mismatches 369; Indels 21; Gaps 4;
 QY 866 TTCTGAGGAGGAGCACTTTATCATGATGAATAAAGCTTCAACCTTCTGAAGGGA 925
 DB 6 TCCGAGGCAAGGCTTTGACTGTCGCTGCTGGAAGGACTTCACTCTCGCTGCG 65
 QY 926 TCAACGGAATTTTTCAGAGACCTCTGACATACCTTACAGTGAAGAGTGGCTGATCTTG 985

DB 66 TGGGTGCAACGCGCTTCCGACACGACGACTACCCCTACGAGAGAGAGTATGACGCTT 125
 QY 986 CCGACGAGCTCGGAATTCCTTGTATGACGAAAGCCCCGACGTTGTATCAAGATACC 1045
 DB 126 GTGACCGTTATGGAATCGTGTATGATGAGAGTCCGGGTGTGGGCATGTCTGGCCC 185
 QY 1046 A---CTACATCCCGAGCTCAGAAATAGACAAACAAATPAAGAAATATAGACA 1102
 DB 186 AGAGCTTCAGCAACGCGTCTCTGACACACACCTTAAGGTATGAGAGATGTTCCGA 245
 QY 1103 GACACAAAGAACATCCCAAGTGTATCATGTGAGTGTGCGAAGCAAGATCCAAACC 1162
 DB 246 GGGACAAAGAACACCCGCGGCTGTGTATGTGTGCGTGGCAATGAGCCCTTCTCTCC 305
 QY 1163 ATCAAGACGCGAGGAGTTCTTCAAGCCCTTTATAGACTGCCAATGAATGATGAA 1222
 DB 306 TGGAAAGGCTGTTTACTACTTCATGATCTGATCCGACACCAAAAGCCTTGACCCCT 365
 QY 1223 CACGCCCCGTTGTCATGTGAGCATGATGAGACGACACGAGAGAAACAGAGAGTGG 1282
 DB 366 CCGGCCCCGTGACCTTCTGTACCA-----GTTCAAGCTATGA---AAAGACCTGG 413
 QY 1283 CGCTGAAGTACTTGCACATGCTGTGTGAACAGTACTACGCTGTATCATCTATCAG 1342
 DB 414 GGGTCCCGTATGTGACGATGATCTGTGAACGTTACTACTCTGTATCATGACTAG 473
 QY 1343 GAAGATTAAGAAAGAGCTTCAAGCTCTGAAAAAACAATAGAAAGCTTATGACAGGC 1402
 DB 474 GTCAATGAGAGGTGATTCAGCTGACAGCTGCCACCCAGTTTGAAGTGGAGTGAAGCCT 533
 QY 1403 ACAGAAAGCCATCTTTCACAGAAATTTGTCGCGAGCCGATAGCTGGCATCCATCAG 1462
 DB 534 ACAGAAAGCAATTAATCAAGAGAGAGTACGGGAGCAAAACATCAAGGTTTACAGAG 593
 QY 1463 ATCCACCTCAATTTCTCCGAAGAGTACCAAGCAGAGCTGTTGAAAAGAGATCAGGC 1522
 DB 594 ATCCACCTCAATTTCTCCGAAGAGTACCAAGAGCCTGTTCAAGCAGTATCATGTGA 653
 QY 1523 TCCTTTTGA---AAAGCTACATCATGGAACAACAGTGTGGCTTTGCAATT 1576
 DB 654 TTCTGTATCAAAAACGCAAAAGATATGTGTGGAAGTCACTGTGAAATTTTGGCGAAT 713
 QY 1577 TTAAGCTCCTCGAATGTGAGAAAGCCATCTCAACACAGAGGTGTTTTCACAAAG 1636
 DB 714 TTTATGCCAGCAATCACTCAAGAGGCCATAGGGAATGAAAAGGATCTTCACTGCAC 773
 QY 1637 ACAGACACCCAAACTGTTGCTCATGTACTGAGA 1671
 DB 774 AGAGACAAACAAAAGTGCAGATTCCTGTCGA 808

RESULT 10

BX363460 1051 bp mRNA linear EST 08-Apr-2004
 LOCUS BX363460 Homo sapiens B CELLS (RAMOS CELL LINE) COR 25-NORMALIZED
 DEFINITION Homo sapiens cDNA clone CSDDL001VM21 5-PRIME, mRNA sequence.
 ACCESSION BX363460
 VERSION BX363460.2 GI:46291531

KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1051)
 Li,W.B., Gruber,C., Jeesee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On May 5, 2003 this sequence version replaced gi:30372720.
 Contact: Genoscope - Centre National de Sequencage
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime

/note=vector: pcwv-sport6/s1ste.1: ECDRV; site 2: Not I. Subtracted with brain, liver, lung, kidney and muscle. Directionally cloned. Priming method: oligo-dT. Average insert size: 1800 bp. Library amplification: 26,000 fold. Kristi A. Eglund, James J. Vincent, Robert Strausberg, Bungkok Lee & Ira Pastan: Discovery of new breast cancer genes encoding membrane and secreted proteins.

CA489317	CA489317	900 bp	mRNA	linear	EST 14-NOV-2002			
LOCUS	ACENCCOURT10810259	MAECL	Homo sapiens	CDNA clone	IMAGE:6721700 5',			
DEFINITION	mRNA sequence.							
ACCESSION	CA489317							
VERSION	CA489317.1	GI:24952108						
KEYWORDS	EST.							
SOURCE	Homo sapiens (human)							


```
Oy 1272 AAGAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGGCTGGTA 1331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 415 CGCAGACCTGGGGGGCCCCCGTACGTGACGTGATTTGTGTGAACAGTTACTTATCCTGGTA 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1332 CATCTATCAGGGAGAGATAGAAAGAACTTCAAGCTTGTGAAAAAGACATAGAAAGCT 1391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 TCATGACTACGGGCATCTGAGAGTGATTCAGCTGCAGCTGACTAGCCAGTTGAGAACTG 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1392 CTATGCCAAGGCACAGAAAGCCCATCTTTGTCAAGAAATTCGGTGGCGAAGCGATAGCTGG 1451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 GTATTAAGATGTACCAAGAGCCAAATTATCCAGAGCGAGTATGGACAGAGCGCGTCTCGGG 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1452 CATCCACTACGATCCACCTCAAAATGTTCTCCGAAAGTACCAAGCAGAGCTGTTGAAA 1511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 GCTTCATAGAGATCCACCTCGCATGTTCAGTGAGAGATGCCAGACAGCTCTCTCGAGAA 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1512 GACGATCAGGCTCCTTTGAAAAAAGA 1538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 TATCATTTGATCCTGGATGAGAAACGA 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: January 24, 2005, 07:30:31
Job time : 5577 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 05:57:45 ; Search time 100 Seconds
(without alignments)
3239.359 Million cell updates/sec

Title: US-09-936-759-6
Perfect score: 3001
Sequence: 1 MVRQNRKKRIFLLINGVNM.....TRDQPKLVAVHRLRMSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 02: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2999	99.9	563	Q9X0F2	Q9X0F2 thermotoga
2	1011	33.7	570	Q9YU11	Q9YU11 enfilobus
3	977.5	32.6	599	Q8XP19	Q8XP19 clostridium
4	972.5	32.4	599	Q8YV14	Q8YV14 clostridium
5	929	31.0	602	Q9AF14	Q9AF14 staphylococ
6	919.5	30.6	598	Q9AHJ8	Q9AHJ8 lactobacilli
7	906	30.2	670	Q6NL66	Q6NL66 drosophila
8	905	30.2	656	AAG93746	AAG93746 drosophila
9	905	30.2	656	Q9Y8R0	Q9Y8R0 drosophila
10	905	30.2	670	Q8WMB7	Q8WMB7 drosophila
11	902.5	30.1	648	BGLR_MOUSE	P1265 mus musculu
12	899.5	30.0	648	Q6IR10	Q6IR10 mus musculu
13	898	29.9	651	BGLR_FELCA	Q97524 felis silve
14	894.5	29.8	648	BGLR_RAT	P06760 ratius norv
15	888	29.6	651	BGLR_CANPA	O18835 canis famli
16	882	29.4	603	Q93YV4	Q93YV4 arabidopsis
17	880	29.3	603	Q8FHA3	Q8FHA3 escherichia
18	879	29.3	603	BGLR_ECOLI	P08804 escherichia
19	879	29.3	603	Q9AHJ4	Q9AHJ4 escherichia
20	872.5	29.1	648	BGLR_CERAE	Q77695 cercopithec
21	865.5	28.8	651	BGLR_HUMAN	P08236 homo sapien
22	865.5	28.8	651	AAG96851	AAG96851 homo sapi
23	862	28.7	603	Q9SQJ2	Q9SQJ2 caenorhadi
24	861.5	28.7	603	Q6W7J7	Q6W7J7 rumiococcu
25	861.5	28.7	603	AAG76046	AAG76046 rumiococ
26	856.5	28.5	808	Q7P1J3	Q7P1J3 ratius norv
27	841.5	28.0	686	Q9Y9T3	Q9Y9T3 drosophila
28	826.5	27.5	599	Q8B6A6	Q8B6A6 streptococc
29	821.5	27.4	599	Q8B0N2	Q8B0N2 streptococc
30	811	27.0	567	Q7UCB6	Q7UCB6 shigella fl
31	809	27.0	567	Q83RC5	Q83RC5 shigella fl

32	728	24.3	660	2	Q7P2E2	Q7P2E2 anopheles g
33	691	23.0	368	2	Q8X671	Q8X671 escherichia
34	691	23.0	370	2	Q7ADL5	Q7ADL5 escherichia
35	531	17.7	459	2	Q8FMX0	Q8FMX0 corynebacte
36	487	16.2	755	2	Q92XF7	Q92XF7 rhizobium m
37	486.5	16.2	745	2	Q93IM0	Q93IM0 thermoaer
38	482	16.1	716	1	BGAL_THERU	P26257 thermoaer
39	482	16.1	716	1	CA316462	CA316462 thermoaer
40	481.5	16.0	998	1	BGAL_LACIA	Q48727 lactococcus
41	477.5	15.9	998	2	Q8VPU3	Q8VPU3 lactococcus
42	476.5	15.9	1085	2	O85250	O85250 thermotoga
43	472.5	15.7	996	2	O87523	O87523 lactococcus
44	460	15.3	743	1	BGAL_THERT	P77989 thermoaer
45	452.5	15.1	1084	1	BGAL_THERA	Q56307 thermotoga

ALIGNMENTS

RESULT 1					
ID	Q9X0F2	PRELIMINARY;	PRT;	563 AA.	
AC	Q9X0F2				
DT	01-NOV-1999	(TREMBlrel. 12, Created)			
DT	01-NOV-1999	(TREMBlrel. 12, Last sequence update)			
DT	01-OCT-2003	(TREMBlrel. 25, Last annotation update)			
DE	Beta-glucuronidase.				
GN	OrderedLocusNames=TW1062;				
OS	Thermotoga maritima.				
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.				
OX	NCBI_TaxID=2336;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=MSB8 / DSM 3109 / ATCC 43589;				
RX	MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;				
RA	Nelson K.E., Claverton R.A., Gill S.R., Gwin M.L., Dodson R.J.,				
RA	Halt D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,				
RA	McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,				
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,				
RA	Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,				
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;				
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from				
RT	genome sequence of Thermotoga maritima.";				
RL	Nature 398:323-329(1999).				
RM	EMBL; AE001766; AAD56143.1; -				
DR	PIR; A72300; A72300.				
DR	HSSP; P08236; 1BHG.				
DR	TIGR; TM1062; -				
DR	GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.				
DR	InterPro; IPR008979; Gal_bind_like.				
DR	InterPro; IPR006101; Glyco_hydro_2.				
DR	InterPro; IPR006102; Glyco_hydro_2ig.				
DR	InterPro; IPR006104; Glyco_hydro_2SB.				
DR	InterPro; IPR006103; Glyco_hydro_2TIM.				
DR	Pfam; PF00703; Glyco_hydro_2; 1.				
DR	Pfam; PF02836; Glyco_hydro_2_C; 1.				
DR	Pfam; PF02837; Glyco_hydro_2_N; 1.				
DR	PRINTS; PR00132; GLHYDRLASE2.				
DR	PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.				
KW	Complete proteome.				
SEQ	SEQUENCE 563 AA; 65682 MW; 98C030B75D33B6C1 CRC64;				
Query Match 99.9%; Score 2999; DB 2; Length 563;					
Best Local Similarity 99.8%; Pred. No. 4.1e-190;					
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
Qy	1	MVRQNRKKRIFLLINGVNMILEVTSKRPPIAVPSSNMEQYDLCYEGSPFTYKTFYYPK 60			
Db	1	MVRQNRKKRIFLLINGVNMILEVTSKRPPIAVPSSNMEQYDLCYEGSPFTYKTFYYPK 60			
Qy	61	XLSDKHRLVPAANTQCEVFLNGEKVGENHIEYLPEVDVTGKVGSENLRAVVENRL 120			

```

Db      61 ELISQKHRLRYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGKVKSGENELRVVENRL 120
Qy      121 KVGFPSPKVPDSGTHTVGFFGSPPPANPDFPPYGGIIRPVLIETDHAHILDIWDTSES 180
Db      121 KVGFPSPKVPDSGTHTVGFFGSPPPANPDFPPYGGIIRPVLIETDHAHILDIWDTSES 180
Qy      181 BEKKLGKVKVIEVSEEAQVGOEMTIKLGEEKIKRTSNRPVGEFFILENARFMSLEDPY 240
Db      181 BEKKLGKVKVIEVSEEAQVGOEMTIKLGEEKIKRTSNRPVGEFFILENARFMSLEDPY 240
Qy      241 LYLPLKVELEKEDYTLIDIGRTISWDEKRLYLNGKRVFLKGGKEEFPVLQGGTFYPLMI 300
Db      241 LYLPLKVELEKEDYTLIDIGRTISWDEKRLYLNGKRVFLKGGKEEFPVLQGGTFYPLMI 300
Qy      301 KDFNLKKNINSFRTSHYPPSEEWLADRLGILVIDEAPVIGITRYHNPETOKIAED 360
Db      301 KDFNLKKNINSFRTSHYPPSEEWLADRLGILVIDEAPVIGITRYHNPETOKIAED 360
Qy      361 NIRMIDRKHNPVSIVMSVANEPESSNHPDAEGFFKALYETANEMDRTRPVVMSMDAP 420
Db      361 NIRMIDRKHNPVSIVMSVANEPESSNHPDAEGFFKALYETANEMDRTRPVVMSMDAP 420
Qy      421 DERTDVALKYPDVCNRYVYGYGRIEGLQALEKDIEDYARHKKPIFYTERGAD 480
Db      421 DERTDVALKYPDVCNRYVYGYGRIEGLQALEKDIEDYARHKKPIFYTERGAD 480
Qy      481 AIAGHYDPQWSESEVQAEIYEKIRLLKKDYIIGHVAFADPKTPQVVRPILNHK 540
Db      481 AIAGHYDPQWSESEVQAEIYEKIRLLKKDYIIGHVAFADPKTPQVVRPILNHK 540
Qy      541 GVFTTRDQPKLVAVHLRRLMSEV 563
Db      541 GVFTTRDQPKLVAVHLRRLMSEV 563

RESULT 2
Q97U11 PRELIMINARY; PRT; 570 AA.
ID      Q97U11
AC      01-OCT-2001 (Tremblrel. 18, Created)
DT      01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      Beta-glucuronidase (Gusb) (EC 3.2.1.31).
GN      Name=gusb; OrderedLocuNames=SSO3036;
OS      Sulfolobus solfataricus.
OC      Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC      Sulfolobus.
NCBI_TaxID=2287;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 35092 / DSM 1617 / P2;
RX      MEDLINE=2133296; PubMed=11427726;
RA      She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA      Aweyer M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
RA      De Moers A., Eruso G., Fletcher C., Gordon P.M.K.,
RA      Heikamp-de Jong I., Jelfries A.C., Kozera C.J., Medina N., Peng X.,
RA      Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA      Charlbois R.L., Doolittle W.F., Duquet M., Gaasterland T.,
RA      Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT      "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RT      Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR      EMBL; AB006894; AAK4318.1; -.
DR      FTR; C90485; C90485.
DR      HSSP; P08236; 1BHG.
DR      GO; GO:0004566; F:beta-glucuronidase activity; IEA.
DR      GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR      InterPro; IPR008979; Gal_bind_1like.
DR      InterPro; IPR006101; Glyco_hydro_2.
DR      InterPro; IPR006102; Glyco_hydro_21g.
DR      InterPro; IPR006104; Glyco_hydro_25B.
DR      InterPro; IPR006103; Glyco_hydro_21m.
DR      Pfam; PF00703; Glyco_hydro_2; 1.
DR      Pfam; PF02836; Glyco_hydro_2_C; 1.

```

```

DR      Pfam; PF02837; Glyco_hydro_2_N; 1.
DR      PRINTS; PR00132; GLHYDLASE2.
KW      Complete proteome; Glycosidase; Hydrolase.
SQ      SEQUENCE 570 AA; 66795 MW; DEB2FEC805AF189 CRC64;

Query Match      33.7%; Score 1011; DB 2; Length 570;
Best Local Similarity 39.2%; Pred. No. 1.7e-58;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

Qy      15 LNVGMNLEVTSSKORP-----IAVGSNNQYQDLCYBGPFTYTTTFYVPK 60
Db      11 LGQFKKFKINENTGENNGYKGLSEBDIITYASNNQPKKDQSGIAMYKDLFVSN 70
Qy      61 XLSQKHRLRYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGKVKSGENELRVVENRL 120
Db      71 DNGRKAMWVFEGAGYITKLMINGEYGTGHEGFTQFKPIKLV-----NEFKIV---V 123
Qy      121 KVGFPSPKVPDSGTHTVGFFGSPPPAN-----FDPPYGGIIRPVLIETDHAHILDT 174
Db      124 KIDNTSPY-----NLPPARDLNNAFDFPNYGGIIRPVYIEFVDECHVEDIT 171
Qy      175 VDTSESEPEKKLGKVKVIEVSEEAQVGOEMTIKLGEEKIKRTSNRPVGEFFILEN 230
Db      172 VYF-----KSYGHLKVEI-LSECNQRFLRFLVDKEGRVILNESSNEVEKD--VNN 222
Qy      231 ARFWSLEDPYLYPLKVELB-----KDEYTLIDIGRTISWDEKRLYLNGKRVFLKGGKE 285
Db      223 VIPWSPNDPYLYTLIVEMVVGNLKDSYVERIGFVDVEYDGIYNGKRPFLKGGGRHE 282
Qy      286 EFPVLQGGTFYPLMIDPFLKKNINSFRTSHYPPSEEWLADRLGILVIDEAP--HV 343
Db      283 DPFILGFTYGAVALVDYFLMKLGANSFRTSHYPSNHLADLADMGVLILEPFLCS 342
Qy      344 GITRYHNPETOKI-----ABDNIRMDRKHNPVSIVMSVANEPESSNHPDAEGF 394
Db      343 NISRVMSQSEIAKMGFDVYKFEKVRTIEMKIRQHNKRSVIMYSVMEPPDIRVAEF 402
Qy      395 FKALYETANEMDRTRPVVMSVANEPESSNHPDAEGF-----HRSVRDLAEYDVVISLNYHGYWTMGDIDSCV 454
Db      403 IREVEFLFKSLDSRRVTPAS-----HRSVRDLAEYDVVISLNYHGYWTMGDIDSCV 457
Qy      455 QALEKDIIEELVARRH-RKPIFYTERFGADAGIHYDPQWSESEVQAEIYEKIRLLKKD 513
Db      458 KVALEIEELIHNKKFPKPIIITFGADAIYGHASPPQWSESEVQSEMKRKITALEKXD 517
Qy      514 YIIGHVAFADPKTPQVVRPILNHKGVFTTRDQPKLVAVHLRLM 560
Db      518 YIVGFHWNFADFRTPQNSRTILNRKGIFTRDQPKLAAYVEELF 564

RESULT 3
O8XP19 PRELIMINARY; PRT; 599 AA.
ID      O8XP19
AC      01-MAR-2002 (Tremblrel. 20, Created)
DT      01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      Beta-glucuronidase.
GN      Name=gubR; OrderedLocuNames=CPE0147;
OS      Clostridium perfringens.
OC      Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC      Clostridium.
NCBI_TaxID=1502;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=13;
RX      MEDLINE=21664373; PubMed=11792842;
RA      Shimizu T., Ohtani K., Hirakawa H.,
RA      Shiba T., Ogasawara N., Hattori M., Kubara S., Hayashi H.;
RT      "Complete genome sequence of Clostridium perfringens, an anaerobic
RT      flesh-eater."
RT      Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR      EMBL; AP003185; BAB79853.1; -.

```

Query Match	32.4%	Score 972.5	DB 2	Length 599
Best Local Similarity	36.7%	Pred. No. 6.5e-56		
Matches 226	Conservative 102	Mismatches 211	Indels 77	Gaps 16
OC	Beta-glucuronidase.			
OC	Clostridium perfringens.			
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;			
OC	Clostridium.			
OX	NCBI_TaxID=1502;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=13;			
RX	MEDLINE=21945361; PubMed=11948145;			
RA	Briolat V., Reyseet G.;			
RT	"Identification of Clostridium perfringens genes involved in the			
RT	adaptive response to oxidative stress."			
RL	J. Bacteriol. 184:2333-2343 (2002).			
DR	EMBL; AJ420784; CAD12654.1; -.			
DR	HSSP; P08236; 1BHG.			
DR	GO; GO:0004553; P:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.			
DR	GO; GO:0005975; P:catabolism; IEA.			
DR	InterPro; IPR008979; Gal_bind_like.			
DR	InterPro; IPR006101; Glyco_hydro_2.			
DR	InterPro; IPR006102; Glyco_hydro_21g.			
DR	InterPro; IPR006104; Glyco_hydro_2SB.			
DR	InterPro; IPR006103; Glyco_hydro_2TLM.			
DR	Pfam; PF00703; Glyco_hydro_2_1.			
DR	Pfam; PF02836; Glyco_hydro_2_C; 1.			
DR	Pfam; PF02837; Glyco_hydro_2_N; 1.			
DR	PRINTS; PR00132; GLHYDRLASE2.			
DR	PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.			
DR	PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; UNKNOWN 1.			
DR	SEQUENCE 599 AA; 68715 MW; BBB5864AAAEPA8 CRC64;			
QY	1	MVRPQAKKRRKIRILLINGVNLVETSKSR-----PLAYGSMNE--QYDOLC 44		
DB	1	MLYPIITTESRQLIDLSGIWFKFLNEGGLTEELSKADLEDTTEMAVPSYNDLVESQEV 60		
QY	45	YEEGPFYKTFYVPYKKSIOGHRILYPAVAVNTDCEVLNGEKVGENHIEVLPFEVDVTGK 104		
DB	61	DHVGWVYERKFTLPKILLNERIYLRGSAIHEKAVYLINGELVHNGGTFPEAEIINDL 120		
QY	105	VKSGENELRVVENRLKVGCFSPKVPDPSGTHVTFGGSFP-----PANEDFPFYG 155		
DB	121	LVSQGNRLTVAVNN-----IIDETTLPVGLVAVEVDGKVIKNSVNFDFNYAG 170		
QY	156	IIRPLLEFTHARILDIWDTSSEBEKLGKVKVKEVSEAVGO-EWTKLGEBEKK 214		
DB	171	IHRPKIYTPPKSYIIBITITVDKKN--GYVNEV---QAVGCKNKVITIDEENN 223		
QY	215	IRTSNRFVEGEFILLNARFMSLEDPVYLKVELKDEYTLD-----IGIRTSIWDCKRL 269		
DB	224	IVABEGKQGLTINNHLMEPMNAVYLKLVLELLDDEBITDYFEEFGVRYLVEVKGKF 283		
QY	270	YLANGPVFLKFGFGHEEPVLGQGTFFPLMKIDFNILKWINANSFRTSHVPYSEHWDLA 329		
DB	284	LINNPFYKFGFGHEDSYVNGRINAINIKDPMKMKIGANSFRISHPIPYSEIMWRLA 343		
QY	330	DLGLILVDEAPVGVITRHYN-----PE-----TOKIAEDNIRRMIDRHN 371		
DB	344	DREGIVVIDETPAVGL--HLNFMATFGGDPAPKRDYTWKEIGTEKAHERILRELVSBDKN 400		
QY	372	HPSYIAMSVAPEPSNHPDAEGFKALYETANEMD-RTRPVVWS--WMDAPDE-RITDVA 428		
DB	401	HPCAVMSVAPEPSDSEGAKEYEPILKLTKEIDPKRPVAVTVYIAMSPPDRCKGVDIV 460		
QY	429	LKFDPIVCVNNYMYIYQGIIEEGLOALEKIEELTAARHK-PIFTEFGADALGIHY 487		
DB	461	----DVLCNRYGYMYAAGDLBEAKRMDELKMGWEERCKPTIMPTEYGCADTVAGLHD 516		
QY	488	DPQMFSEERYQAEIVLVEKTIIRLLKKDYIIGTHVAFADFPTQNVRRPILNKHGVFTDR 547		
DB	517	TVPMVFTEERYVEYKKNHHEVMDCKNFVEQVWNPADPFTSGIITVQCNKKGIIFRRR 576		

Qy 548 OPRVAHVLRRLMSEV 563
Db 577 KPRMLAHLRERWTNI 592

RESULT 5

Q9AFA2 PRELIMINARY; PRT; 602 AA.
AC Q9AFA2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Beta-glucuronidase.
OS Staphylococcus sp. RUH1.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=156489;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RLH1;
RA Jefferson R.A., Keese P.K.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RC SEQUENCE FROM N.A.
RA Kilian A.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
DR HSBP; P08236; 18HG.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2ig.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TLM.
DR Pfam; PF00703; Glyco_hydro_2_1.
DR Pfam; PF02836; Glyco_hydro_2_C_1.
DR Pfam; PF02837; Glyco_hydro_2_N_1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
SQ SEQUENCE 602 AA; 68701 MW; 312AFCD1634D577 CRC64;

Query Match 31.0%; Score 929; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 4.9e-53;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

Qy 1 MVRPQNRKRPILLNGVWNLV-----TSKDRPIAVPGSMNE--QYODLC 44
Db 1 MLYPINTETRGVFDLNGVWNLKLDYKGGLBEKWEKSLDTISMAVSSYNDIGVTKEIR 60
Qy 45 YEEBPFTYKTFYPPKLSQKHILLYFAAVNTDCEVLNGEKGEHIEYLPEVDVTGK 104
Db 61 NHIVYWEREFYPAVLKQRIYLRFGSATHKAIYVNNELVYEHKGFLEFAEAINNS 120
Qy 105 VKSGENELRVVENRLKVGFPFKVPSGTHTVGFGS-----PPANDPPPYG 154
Db 121 LRDMNVTVAVDIL-----DDSTLPVGLYERHEEGIGKIRKNKPNDFPNYA 170
Qy 155 GIIPVLIETFDHARILDIWDTSESPEKKLGKVKVIESEBANGQENTIKLG--EE 212
Db 171 GLHNPVIXITYTPFYVEDISVTDENGP---TGIVTVYDFQ---GKAETVKSVDDE 223
Qy 213 KKIIRTSRPFVEGEIILNARFWSLEDPLVPLKYLEKDEYLDI-----GIRTSIDEX 267
Db 224 GKVAASTEGSGNVEIENVILMEPLNTLYQIKVELVNDGLTIDVVEBPGRVTEVNDG 283
Qy 268 RLYLNGKRVPLKGFGEKEEFPVLGQGTFFYLMIKDFNLKMINANSPRTSHYVSEWLD 327
Db 284 KFLINNKPFYFKGKGKEDPTINGRGFNEASNWDFNIIKMGANGSRTIHYVSEELMR 343
Qy 328 LADRLGLIVIDEAPHVIGITRYHTN-----PETOKIA-----EDNIRMRIDR 368

Db 344 LADREGLVVIDETPAVG---HLNFMAATGIGGSGSERVSTWERTIRTEHHQDVLREIVSR 400
Qy 369 HKHNPVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVWVSMMDADDERRDV 427
Db 401 DKHNPVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVWVSMMDADDERRDV 427
Qy 428 ALKIFDIVCNRRYYGWIYQRIEGLQALEKDIIELYARHR-----KPIFTEFGADA 483
Db 459 VAEIIVIALNRXNGWYFDCGDLBAKVHLR---QEFHANNKQCPKPIMITTEYGDVTA 515
Qy 484 GIHDPQMSSESOAELVKTIRLLKDYITIGTHVMAADPTQPNRRPILNKGVF 543
Db 516 GFHDIDVMPTEBYQVEYYOANHVFDEFENFVGOAMNADPATSGVNRVQGNKGVF 575
Qy 544 TRDRQPLVAHVLRRLMSEV 563
Db 576 TRDRKPLAHLRERWTNI 595

RESULT 6

Q9AHJ8 PRELIMINARY; PRT; 598 AA.
AC Q9AHJ8;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Beta-glucuronidase.
OS Name=guaA;
GN Lactobacillus gasseri.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1596;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ADH;
RX MEDLINE=21141841; PubMed=11229918;
RA Russell W.M., Kleenhammer T.R.;
RT "Identification and cloning of guaA, encoding a new beta-glucuronidase
from Lactobacillus gasseri ADH."
RL Appl. Environ. Microbiol. 67:1253-1261 (2001).
DR EMBL; AF05888; AAK07836.1; .
DR HSBP; P08236; 18HG.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2ig.
DR InterPro; IPR006103; Glyco_hydro_2SB.
DR InterPro; IPR006104; Glyco_hydro_2TLM.
DR Pfam; PF00703; Glyco_hydro_2_1.
DR Pfam; PF02836; Glyco_hydro_2_C_1.
DR Pfam; PF02837; Glyco_hydro_2_N_1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
SQ SEQUENCE 598 AA; 69762 MW; 5358F06B2DB87D CRC64;

Query Match 30.6%; Score 919.5; DB 2; Length 598;
Best Local Similarity 34.1%; Pred. No. 2.1e-52;
Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;

Qy 4 PORNRKRPILLNGVWNLV-----SKDRP-----IIVPGSMNE--QYODLCYE 47
Db 7 PIONKRPFTLNGWQVETDPSNVGLDEGMNKLDPDEPMVPGTFALTTKRKRKYT 66
Qy 48 GPFTYKTFYVPKLSQKHILLYFAAVNTDCEVLNGEKVGENHIEYLPEVDVTGKYS 107
Db 67 GDFWYQKDFRISPLKKELEYIRFGSVTHRAKVFINGHEVQHBEGFLPFQVKINNYIN 126
Qy 108 GE-NELRVVENRLKVGFP---SKVPSGTHTVGFPFGSPPPANDFPFYGGIIRPVLE 163
Db 127 DQTNRTVTVLVNNELEKAIPLGTETELDNGQ-----KLAQPYDPFNYSGIMENVMVL 179

QY 164 FTDHARLIDWVDTSESEPEKLGKVKYKIEVSEAVGO-EMTITLGESEKKIRTSNRPV 222
Db 180 ALPQOSITNFKLN-----YQANNKATITTYNIEANNNAEFKVTLPDQKQVACVTSKN 232
QY 223 EGEFFLENNARFMSLEDDPYLPKVELEK-----DEYTLDIGIRITSMDEKRLYLNGKPVF 277
Db 223 TSSLITKNPHLMSPPNDPYSYKIKIEMLEBDGKTVDETDQIGIRYKIVNDKLLNNHPY 292
QY 278 LKGFGEHEEPVLGGGTFFPLMIKDFNLKWINANSFRTSHYPSYSEWMLDLRLGILVI 337
Db 223 LKGFGEHEEPVLGGGTFFPLMIKDFNLKWINANSFRTSHYPSYSEWMLDLRLGILVI 352
QY 338 DEAPHVGIITRYHYN-----PEFOKIAEDNIRIMIDHKKHPSVIMMS 379
Db 333 DEAPHVGLMRSTTNLNTNSNOSHFPASKVTPBELKVEHOBIEKENDIDQHPSPVIAWS 412
QY 380 VANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVMSMDADERTDVALKYPIVCV 438
Db 413 LENEBESTTQESYDFKQIFAFARAKLDPQNRPYTGLVWGSQPK--VDKHLPLCDPVCIN 470
QY 439 RYGYHYIYOG-RIEBGLQALEKQIEBLY-ARRHKPIFYTEFGADALAGIHYDPQPMFSEE 496
Db 471 RYGYHYVAGGPEIVNAKKMLDELDQWQYLKLNKPFVFEFGADTLSSSHRLPDEWMSQ 530
QY 497 YQAEVLEKTIIRLLKKDYIIGHVAFADFKTPQVNRRIPLNHKGVTFRDRQPKLVANHL 556
Db 531 YQAEVLEKTIIRLLKKDYIIGHVAFADFKTPQVNRRIPLNHKGVTFRDRQPKLVANHL 590
QY 557 RRLMSEV 563
Db 591 KRWQQL 597

RESULT 7
Q6N166 PRELIMINARY: PRT: 670 AA.
AC Q6N166;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE RE15795P.
GN Name=CG15117;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Carlson J., Chavez C., Friese E., George R., Pacleb J.,
RA Park S., Wan K., Yu C., Rubin G.M., Celniker S.;
RL Submitted (Apr-2004) to the EMBL/Genbank/DBJ databases.
EMBL: BT012475; AAS93746.1; -;
DR InterPro: IPR008979; Gal_bind_1like.
DR InterPro: IPR006101; Glyco_hydro_2.
DR InterPro: IPR006104; Glyco_hydro_21g.
DR InterPro: IPR006103; Glyco_hydro_2SB.
DR InterPro: IPR006025; Pept_M_zn_BS.
DR Pfam: PF00703; Glyco_hydro_2_1.
DR Pfam: PF02836; Glyco_hydro_2_C_1.
DR Pfam: PF02837; Glyco_hydro_2_N_1.
DR PRINTS: PR00132; GLHYDRLASB2.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 670 AA; 77050 MW; 00AB0E67AE1D9E8C CRC64;

Query Match 30.2%; Score 906; DB 2; Length 670;
Best Local Similarity 34.7%; Pred. No. 1.9e-51;
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;

QY 1 MVRPQRNKRIFILINGVNL-----EVTSKDR---PIAVPSNMQ 39
Db 48 MLYPSESETRVRSIDGIMNFRSDQANPTQVRDEWYAKELSKSRPTTPMEVSPASYNDI 107
QY 40 YOD-LCYEGEPPTYKTTFFVPKLSQ-KHIRLYFAVNTDCEVFNGEKVGNHLEYLPF 97
Db 108 TTDNLRDHYGVWYRKFFVRSMKQDRIWLRFSVHYEALVWNGQKVYHGHGILPF 167
QY 98 EVDVTGKYSQ-ENELRVVENRLKVGFP-----SKVPDSGTHVGFSGSPRANFDP 152
Db 168 EAEVTDLSVGENRITVMCDVALIQTVPQRIIEVNDQGMTI-----VQSYTFDFEN 222
QY 153 YGGIRPLVLEPTDARLIDWVDTSESEPEKLGKVKYKIEVSEAVGO-----EMTITK 207
Db 223 YAGHRSVHLVYTPPTFFIEEVAVTNLSK-DATIGEVYSVSVNSAANEADNVLQIQAN 281
QY 208 LGESEKKI---RTSRPFGESEFLENARFW-----SLDDPYLPKVELEK-----DEY 253
Db 282 LYDKGILVANATSDQKLGKLGKLVVVPVFWPVMYLMHSEBGIYQLEIKLATNDELIDVY 341
QY 254 TLIDIGIRITSMDEKRLYLNGKVPFLKGFGEHEEPVLGGGTFFPLMINDNLKWINANS 313
Db 342 RLKVGIRITLMSWSQOFLNGKVPVFRGGRHDSIDIRGGLDNALMVRDNLKWTIGANA 401
QY 314 PRTSHYPSSEWMLDLADRLGILVIDEAPHVGIITRYHYNPEFOKIAEDNIRIMIDHKKHP 373
Db 402 YRTSHYPSSEWMLDLADRLGILVIDEAPHVGIITRYHYNPEFOKIAEDNIRIMIDHKKHP 459
QY 374 SVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVMSMDADERTDVALKYPD 433
Db 460 SVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVMSMDADERTDVALKYPD 515
QY 434 IVCNRYGYWYIYOGRIEGLQALEKQIEBLY-ARRHKPIFYTEFGADALAGIHYDP 490
Db 516 IISFNRYNAMYSNAGRLD---MITQNVIDEALAMNKRYNKPIIMSEYGDITLEGHMDPA 572
QY 491 QMFSEFQAEVLEKTIIRL---LLKKDYIIGHVAFADFKTPQVNRRIPLNHKGVTFRDR 547
Db 573 YWSEEFQAEVLEKTIIRL---LLKKDYIIGHVAFADFKTPQVNRRIPLNHKGVTFRDR 632
QY 548 QPKLVANHLRK 558
Db 633 QPKLVANHLRK 643

RESULT 8
AAS93746 PRELIMINARY: PRT: 670 AA.
AC AAS93746;
DT 23-APR-2004 (TREMBLrel. 27, Created)
DT 23-APR-2004 (TREMBLrel. 27, Last sequence update)
DT 23-APR-2004 (TREMBLrel. 27, Last annotation update)
DE RE15795P.
GN CG15117;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Carlson J., Chavez C., Friese E., George R., Pacleb J.,
RA Park S., Wan K., Yu C., Rubin G.M., Celniker S.;
RL Submitted (Apr-2004) to the EMBL/Genbank/DBJ databases.
EMBL: BT012475; AAS93746.1; -;
SQ SEQUENCE 670 AA; 77050 MW; 00AB0E67AE1D9E8C CRC64;

Query Match 30.2%; Score 906; DB 2; Length 670;
Best Local Similarity 34.7%; Pred. No. 1.9e-51;
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;

1 MVRPQRNKRIFILINGVNL-----EVTSKDR---PIAVPSNMQ 39

Db 48 MLVPESEERREVRSLDGINVFRSDQANPTGVRDEMTAKELSKRPITMPEVPSYNDI 107
 Qy 40 YOD-ICIEGEPPTTYKTTPVPCKLQ-KHRLYPAAVNTDCEVLNGEKVGNHIEYLP 97
 Db 108 TTNDNRDHGVWYRKRFVPRSRWSDKDRIMLRFSGSVHAEVYVWNGKVVHEGHLP 167
 Qy 98 EVDVATGKVGSG-ENELRVVENVRLKVGPP-----SKVDSGHTWNGFGSPFPANPDFP 152
 Db 168 EAEVVDLISYGAENKITWCDNALQTTVPQCRITVEVFDGSMIT-----VQSYTFDFPN 222
 Qy 133 YGGIIRPVILIEFTDARILDIWDTSSEPEKKLGKVKYKLEVSBAVQ-----EMTIK 207
 Db 223 YAGIRSHVILYTPPTFIEVEVETTNLKG-DATICEVPSYVNSGANENADNVLQIQAN 281
 Qy 208 LGESEKCI---RTSRPFVEGEFIEENAFW-----SLSDPILYPLKVELEK-----DEY 253
 Db 282 LYDKGGLVAAVNTSQKGGKQLQVNPVPRWMPYLMHSEBPGYLYOLEIKLMTNDELDEY 341
 Qy 254 TLIDIGRTISMDKRLYNGKPFVFLKGFKEEFVLQGTFFYPLMIDKFNILKWNANS 313
 Db 342 RLKVGIRILSNNSQOPLNGKRVYRGRGHEHDSIRGKLDNALVWDFVLLKWTGANA 401
 Qy 314 FRTSHYPSYSEWLDIADRLGLIVIDEADPHVGTIRYHNPETOKIAEDNIRMRIDRHKHP 373
 Db 402 YRTSHYPSYSESMQPADHGHIMTIDECPSVDTE--NFSOELLGKHKSLLEGILHRDRNH 459
 Qy 374 SVINWSVNPESNPNDAGEFPKLYETANEMDRFRPVVMSMDAPERRDVALKFPD 433
 Db 460 SVVMSINPERITGVSADSTFELVANFTSLDKRPTTAIAV-----SNQODKAGSLD 515
 Qy 434 TVCVNRVYVWYIYOGRIEGLQALEKIDIELYA---RHRKPIFVTEFGADAIAGIHYDP 490
 Db 516 IISFRVYVWYVSNGRID---MIGNVIDEAIAANKRKNTIMSSEYCADLBLEHNOQA 572
 Qy 491 QMFSEYVAAELVEKIRL---LLKKDYIIGTHVAFADPCKTPONVRPILNHKCVFTDR 547
 Db 573 YVWSEEFQTEVFSHRFKAFDELRRKGMFIEGFVWVNFADFKTAQGYTRVGNKKGVFRAR 632
 Qy 548 QPKLVAVLRR 558
 Db 633 QPKAAHLRR 643
 RESULT 9
 Q9V8R0 PRELIMINARY; PRT; 656 AA.
 ID Q9V8R0; ORF007; 656 AA.
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
 DE CGJ5117-PA (LD04718p).
 GN ORFNames=CGJ5117.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Archipoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidae; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dunin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howard T.J., Wei M.H., Ibegwan C.,
 RA Jatali W., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Kimball B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser K.D., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palaret K., Remington K., Saunders R.D., Scheeler F., Shen R.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkac R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodruff D., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svrtkac R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
 RL melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Celniker S.E., Bergman C.M., Kronmiller B., Carlson J., Svrtkac R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 RL a genomic perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Croebly M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker U.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Beutenkourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RL systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426071; PubMed=12537574;
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426072; PubMed=12537575;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426073; PubMed=12537576;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426074; PubMed=12537577;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426075; PubMed=12537578;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426076; PubMed=12537579;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [11]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426077; PubMed=12537580;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426078; PubMed=12537581;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [13]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426079; PubMed=12537582;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426080; PubMed=12537583;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [15]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426081; PubMed=12537584;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [16]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426082; PubMed=12537585;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [17]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426083; PubMed=12537586;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [18]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426084; PubMed=12537587;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [19]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426085; PubMed=12537588;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [20]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426086; PubMed=12537589;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [21]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426087; PubMed=12537590;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [22]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426088; PubMed=12537591;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [23]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426089; PubMed=12537592;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [24]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426090; PubMed=12537593;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [25]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426091; PubMed=12537594;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [26]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426092; PubMed=12537595;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [27]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426093; PubMed=12537596;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [28]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426094; PubMed=12537597;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [29]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426095; PubMed=12537598;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [30]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426096; PubMed=12537599;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [31]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426097; PubMed=12537600;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [32]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426098; PubMed=12537601;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [33]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426099; PubMed=12537602;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [34]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426100; PubMed=12537603;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [35]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426101; PubMed=12537604;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [36]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426102; PubMed=12537605;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [37]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426103; PubMed=12537606;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [38]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426104; PubMed=12537607;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [39]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426105; PubMed=12537608;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [40]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426106; PubMed=12537609;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [41]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426107; PubMed=12537610;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [42]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426108; PubMed=12537611;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [43]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426109; PubMed=12537612;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [44]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426110; PubMed=12537613;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [45]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426111; PubMed=12537614;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [46]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426112; PubMed=12537615;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [47]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426113; PubMed=12537616;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [48]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426114; PubMed=12537617;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [49]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426115; PubMed=12537618;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [50]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426116; PubMed=12537619;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [51]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426117; PubMed=12537620;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [52]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426118; PubMed=12537621;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [53]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426119; PubMed=12537622;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [54]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426120; PubMed=12537623;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [55]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426121; PubMed=12537624;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [56]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426122; PubMed=12537625;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [57]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426123; PubMed=12537626;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [58]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426124; PubMed=12537627;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [59]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426125; PubMed=12537628;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [60]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426126; PubMed=12537629;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [61]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426127; PubMed=12537630;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [62]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426128; PubMed=12537631;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [63]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426129; PubMed=12537632;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [64]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426130; PubMed=12537633;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [65]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426131; PubMed=12537634;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [66]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426132; PubMed=12537635;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [67]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426133; PubMed=12537636;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [68]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426134; PubMed=12537637;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [69]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426135; PubMed=12537638;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [70]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426136; PubMed=12537639;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [71]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426137; PubMed=12537640;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [72]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426138; PubMed=12537641;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [73]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426139; PubMed=12537642;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [74]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426140; PubMed=12537643;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [75]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426141; PubMed=12537644;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [76]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426142; PubMed=12537645;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [77]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426143; PubMed=12537646;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [78]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426144; PubMed=12537647;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [79]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426145; PubMed=12537648;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [80]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426146; PubMed=12537649;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [81]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426147; PubMed=12537650;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [82]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426148; PubMed=12537651;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [83]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426149; PubMed=12537652;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [84]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426150; PubMed=12537653;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [85]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426151; PubMed=12537654;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [86]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426152; PubMed=12537655;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [87]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426153; PubMed=12537656;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [88]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426154; PubMed=12537657;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [89]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426155; PubMed=12537658;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [90]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426156; PubMed=12537659;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [91]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426157; PubMed=12537660;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [92]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426158; PubMed=12537661;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [93]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426159; PubMed=12537662;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [94]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426160; PubMed=12537663;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [95]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426161; PubMed=12537664;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [96]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426162; PubMed=12537665;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [97]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426163; PubMed=12537666;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [98]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426164; PubMed=12537667;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [99]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426165; PubMed=12537668;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [100]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426166; PubMed=12537669;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [101]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426167; PubMed=12537670;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [102]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426168; PubMed=12537671;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [103]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426169; PubMed=12537672;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [104]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426170; PubMed=12537673;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [105]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426171; PubMed=12537674;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [106]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426172; PubMed=12537675;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [107]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426173; PubMed=12537676;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [108]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426174; PubMed=12537677;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [109]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426175; PubMed=12537678;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [110]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426176; PubMed=12537679;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [111]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426177; PubMed=12537680;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [112]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426178; PubMed=12537681;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [113]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426179; PubMed=12537682;
 RA FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN [114]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426180; PubMed=12537683;
 RA FlyBase;
 RL Submitted (

RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungali C.J.,
 RA Munoz J., Pacleb J., Paragas V., Park S., Pounanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003797; AAF57602.2; -
 DR EMBL: AY069329; AAL39474.1; -
 DR HSSP: P08936; 1BHG
 DR FLYBase: FBgn0034417; CG15117.
 DR GO: GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro: IPR008979; Gal bind like.
 DR InterPro: IPR006101; Glyco_hydro_2.
 DR InterPro: IPR006104; Glyco_hydro_2ig.
 DR InterPro: IPR006103; Glyco_hydro_2SB.
 DR InterPro: IPR006103; Glyco_hydro_2T1M.
 DR Pfam: PF00703; Glyco_hydro_2; 1.
 DR Pfam: PF02836; Glyco_hydro_2_C; 1.
 DR Pfam: PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS: PR00132; GLHYDRLASEZ.
 DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKOWN 1.
 SQ SEQUENCE 656 AA; 75166 MW; 34168033CBF80A20 CRC64;

Query Match 30.2%; Score 905; DB 2; Length 656;
 Meet Local Similarity 34.7%; Pred. No. 2,1e-51;
 Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;

1 MVRPORKKRFILNGVNL-----EYTSKR---PIAVGSNMQ 39
 34 MLYPSESTREVRSLDGIWNPFRSQDANPTQCVDEWYAKELSKRPITPMFVPSYNDI 93
 40 YQD-LCYEGEPFYKTTFFVPPKLSQ-KHRLYFAVNTDCEVFLNGKVGSENIIEYLPF 97
 94 TTDNLRDHVGTVWDRKFFVPRSWSKDQRIWRFCSVYHEAVVWNGKVHNGHGLPF 153
 98 EVDVTGKAKSG-ENELRVVENRNLKVGFR---SKVPDSGTHVYGFSGPPRANFDFRP 152
 154 EAVVTDLSYGAENRITWCDNALIQTVFPQGRITEVNDGDMIT-----VQSYTFDFPN 208
 153 YGGIRPVLIERFTHARLIDIVWTSSEPEKKLGKVKYKIEVSEAVGQ-----EMTK 207
 209 YAGIRSHVLTPTPTFIEEVEVTNLK-DATVGEVTVSYVNSGASANEADNVLQIQAN 267
 208 LGESEKTI---RTSNRFEGEFIEENARFW-----SLEDPPYLPLKVELEK-----DEY 253
 268 LYDKGILVANATSDQKLGKLGKQVNPVPRWRYLHNSGVLVYQGEIKLATNDELLOY 327
 254 TLDIGIRITSDEKRLYNGKVPFLKGFQKHEEPVLCQGTFFYPLMIKDFNLKINANS 313
 328 RLKVGIRTLSTWNSQOFLNGKRPVYRGRGHEDSDIRGGLDNALMVDFNLKIGANA 387
 314 FRTSHYPSSEWLDLADLGLILVDEAPHYGTRHNPETQKIMEDNIRIMIDKHKNP 373
 388 YRTSHYPSSEWQPADHGLMITIDECPSVDTE--NFQOELIGKSKSLSEQIHHDRNP 445
 374 SVIMVSNPESESNPDAEGEFKALYETANEMDRTPVVMVSMADPDETRDVALKTFD 433
 446 SVVMVSNPESESNPDAEGEFKALYETANEMDRTPVVMVSMADPDETRDVALKTFD 501
 434 IVCNRYRYGWTIYQRIEGLQALEKIDBELYA---RHKKPIFTEFGADALAGIHYDP 490
 502 IISFRRYVAYWNSNARLD---MITQNVIDEALAMKRYKPKIIMSEYADLTLEGHMOFA 558
 491 QMFSEYQALEVEKTIKRL---LTKKDYIIGTHVAFADPKTPONRRPILNHKGVFTDR 547
 559 YWSEEFTEVFSRPFKAFDELKKGWFIIGFVWNPADPKTAQSYTRVGANKGVFTAR 618
 548 QPKLVAVHLRR 558
 619 QPKAAAHLLRK 629

RESULT 10
 Q8MB7 PRELIMINARY; PRT; 670 AA.
 AC Q8MB7;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE CG15117-PB.
 GN ORFNames=CG15117;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.M., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abil J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Duthin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harlin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.V., Wei M.H., Ibegyan C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laevo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X.P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X.P., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson S.M., Nelson K.A., Nixon K., Nussken D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reiner K., Remington K., Saunders R.D., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodgett M., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskaas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
 melanogaster euchromatic genome sequence."
 RN Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,
 RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,

RA Ashburner M., Celisner S.E.:
 RA "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RA a genomics perspective.",
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RL [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2426069; PubMed=12537572;
 RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Battencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RL [5]
 RP SEQUENCE FROM N.A.
 RG FLYBASE;
 RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RL FLYBASE;
 RG EMBL, AE003797; AA068434.1; -.
 RG HSSB; P08236; IBBG.
 DR FLYBASE; FBgn003417; CG15117.
 DR GO; GO:000455; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR008979; Gal bind like.
 DR InterPro; IPR006101; Glyco_hydro_2.
 DR InterPro; IPR006102; Glyco_hydro_2ig.
 DR InterPro; IPR006104; Glyco_hydro_2SB.
 DR InterPro; IPR006103; Glyco_hydro_2TM.
 DR InterPro; IPR006025; Pept_M_zn_BS.
 DR Pfam; PF00703; Glyco_hydro_2; 1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLYDRLASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
 DR SEQUENCE 670 AA; 77036 MW; 7a9P23351E458C90 CRC64;

Query Match	30.2%;	Score 905;	DB 2;	Length 670;
Best Local Similarity	34.7%;	Pred. No. 2.2e-51;		
Matches 212;	Conservative 122;	Mismatches 209;	Indels 68;	Gaps 177;

```

QY 1 MVRFOHKKRIRILLINGVWNL-----EYTSKCR---PIAVPGSANNQ 39
Db 48 MLVYRFBSTREYRSLDGMVFNVRSDQANPTQGVREDWYAKELISKRPTPIWEPVASYNDI 107
QY 40 YQD-LCYEEGFTYKTFEYVFKLSQ-KHILYPAAVNTDCEVFLNGEKVENHIEYLPF 97
Db 108 TTDMLRHWGVGVMWYDRKFEVPRSRMSKDQRIMLRFGSVHYEAUWVINGQKVVXHEMGHLPF 167
QY 98 EVDYTGKYSG-KEXELRVVENRKLKVGFP---SKYVDSGTHVYFGSGSPRPNDFPF 152
Db 168 EAEYTDLLSYAEKRIYVMCDNALIQTVPGGRITFEVNDGAMT-----VQSYTFDFPN 222
QY 153 YGGIIRPVLLFEFDHARILDIWVDTBSSEPEKKLGYKVKILEVSEAVQ-----EMITK 207
Db 223 YAGIHRSEVHLTYTTRTFEEVEVTTNLSK-DATVGEVFSVSVNGSAAENADVLIQIQAN 281
QY 208 LGSEBEKI---RTSNRFVEGFEIENLRFM-----SLDDPYLPLKYLEK-----DEY 253
Db 282 LYDKDGLIVANATSDQKLQGLQVNPYKPMWRYLHMHSBGLIYQLEITGLATNDELIDVY 341
QY 254 TLIDIGITISWDEKRLYLNGKPVVLKGFQKHEEPVVLQGGTFYPLMTKDENLLKWINANS 313
Db 342 RLKVGITLWSNSQGFINGKRPVYFRFGREDDISGKIGDNLAMVDFMLLWIGINA 401

```

```

OY 31 FRSHHPYSEEMULADRLGILVIDEAHPYVIGTRHYNPEQOKAEDNIRMIDRKHNP 373
Db 402 YRSHHPYSEESQOFDEHGMITIDECSVDTE--NFSQELLGKMKSSLEGLIHRDNHP 455
OY 374 SYIMWSVANEPESNHDPDAEGFKALYETANEMDRTPVWWSMMADADERTRDVALKYFD 433
Db 460 SYVMWSIANEPRTGSVADSAPFELVAFNFRSLDKRPTITALIANV-----SNTQDAGKSLD 515
OY 434 IVCWNYRYGMYIYQGIIEEGLOALEKQDELYA--RRKRITPYTEGADAIAGIHYDP 490
Db 516 IISFNKNMYSNAGSLD--MITQNVIDEIAIANNKRNRNKIIMSEYGAJTLBELHNQPA 572
OY 491 QMESBEYOALVEKTIRTL---LLKDYIIGITHVAFAEDFKTPQWVRRLINHKGVFTRDR 547
Db 573 YVMSEEFQREVSRRHKFADELRRKKGWPIGEFVWNNPADFKAQSYTTVGSKKKGVFTRAR 632
OY 548 QPKLVANHYLR 558
Db 633 QPFAAAHLIRK 643

```

RESULT	11
BGLR_MOUSE	
ID_BGLR_MOUSE	STANDARD; PRT; 648 AA

AC P12265; Q61501; Q64473; Q64474;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=Gusb; Synonyms=Gus; Gus-s;
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
X

RN
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88085188; PubMed=2891607;
 RA Gallagher P.M., D'Amore M.A., Lund S.D., Elliott R.W., Pazik J.,
 RA Hohman C., Korthagen T.R., Gaischow R.E.;
 RA "DNA sequence variation within the beta-glucuronidase gene complex
 RT among inbred strains of mice";
 RL Genomics 1:145-152(1987).
 [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=88284700; PubMed=3397060;
RA Gallagher P.M., D'Amore M.A., Lund S.D., Ganschow R.E.;
RT "The complete nucleotide sequence of murine beta-glucuronidase mRNA
and its deduced polypeptide";
RL Genomics 2:215-219 (1988).

RN
RN
SEQUENCE FROM N.A.
MEDLINE=89062453; PubMed=3196706;
P./Amere V., Gallagher P.M., Korfagen T.R., Ganschow R.E.;
RT "Complete sequence and organization of the murine beta-glucuronidase
gene";
RL Biochemistry 27:7131-7140(1988).

KN SEQUENCE FROM N.A.
 RP STRAIN=YBR, and C3H/HeJ; TISSUE=Sperm;
 RC MEDLINE=89384641; Pubmed=2779578;
 RX MAWRZYNAK C.J., GALLAGHER P.M., D'AMORE M.A., CARER J.E., LUND S.D.,
 RA RINDOK E.M., GANSCHKE R.E.;
 RT "DNA determinants of structural and regulatory variation within the
 RT murine beta-glucuronidase gene complex.";
 RL Mol. Cell. Biol. 9:4074-4078(1989).

RN SEQUENCE FROM N.A.
 RP MEDLINE=88216590; PubMed=2835664;
 RA Funderstein B,¹ Leary S.L., Stein J.C., Catterall J.F.,
 RT "Genomic organization and sequence of the Gus- α allele of the
 RL murine beta-glucuronidase gene.";
 RL Cell. Biol. 8:1160-1168(1988).
 CL -1- FUNCTION: Plays an important role in the degradation of dermatan

QY	202	QEMTICKAGEBKRTSNRPFGEFTLEBARW----	SLEDP-YLYPLKY-----	ELEKD	251
Db	256	-OLEVQLLDEBGKVAHAGTGNQGLQVPSANLWPMPLMEHPAWMVLSLEKVTYTTESVTD			314
QY	252	EYTLDIGRTISWDEKRLYLNLKGPFLKSGFGKHEEPVVGQGTFFPLMTIKDFLLKWINA			311
Db	315	YTLPLVGRITVAVYTKSKPLINKPFFYFGQVNHGSDINGKGFDMVLKDFLLKWLGA			374
QY	312	NSFRTSHYPYSEEWLIDLADRLGILIVIDEAPHVGIT-----	RHYNPETOKIAEDN		361
Db	375	NSFRTSHYPYSEEWLQCDRQYIVVIDECPGVIGLPGQFGNESLRLHL-----	EWVBEEL		429
QY	362	IRRMIDRKHNHPSVLMVSNVANEPSNHPDAEEFPFALYLTANEMDRTREYVWVS--	MMDA		419
Db	430	VR-----RDKMHPAAVVMWSVANEPSALKPAAYVFTLLTHTRKALDLTRPVTFVSNAKYDA			485
QY	420	PDETRDVALKCFDVCNVRRYGWTIYOGRIEGLQALEKDIETLAFRRKPFVTFEFGA			479
Db	486	-----DLGAPYVDICVNSYFSTWHDGHLEVIQPOLNSQFENMYKTHOKPLIOSEYGA			539
QY	480	DALAGIHVDPQPMFSEEXQALVEKTRIL--LKDDYIIGTFVMAFADKTPONVRRPIL			537
Db	540	DALPEIHDDPRMFSEEXQKAVLENVHSTLDCKREYVVGELIWMNADPMTQNSPLRVIG			559
QY	538	NHKGVFTDRQPKVAHVLR-RLW	560		
Db	600	NKKGIFTRORPKTSAFILRERYW	623		
RESULT 12					
ID	Q6IR10	PRELIMINARY;	PRT;	648	AA.
AC	Q6IR10;				
DT	05-JUN-2004	(TREMBlrel. 27, Created)			
DT	05-JUN-2004	(TREMBlrel. 27, Last sequence update)			
DT	05-JUN-2004	(TREMBlrel. 27, Last annotation update)			
DE	Beta-glucuronidase.				
GN	Name=Gus;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NM1; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;				
RX	MEDLINE=22388257; PubMed=12477932;				
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,				
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyak S.W.,				
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,				
RA	Kryzhanovskii I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,				
RA	Jones S.J., Marz M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).				
RP	(2)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NM1; TISSUE=Mammary tumor. WAP-Tag model. 5 months old;				
RA	Strausberg R.,				
RA	Submitted (JUN-2004) to the EMBL/GenBank/DBSJ databases.				
RR	EMBL; BC071226; AAH1226.1; -				
RR	Interpro; IPR008979; Gal bind like.				

Db 204 NPDEFNVAGLRPVLYTPTTPTTIDITISTSVNQ-DTGLVDVQYFVGESEHF---QLEV 259
Qy 207 KLGESEKIRTSNRFVEGEFTLENNARFW---SLEDP-YLYPLKYLE-----KDEY 253
Db 260 RLDBEGKRVAGCTGCGRQLQVPRNAHLMWRYLMHBPALVLSLEVTALQTAAGSVDFY 319
Qy 254 TLDIGIRTSMDKRLYLNGKRVPLKGFQKHEEPVLQGTFFYPLMIDPNLLKMINNS 313
Db 320 TLPVGIRTVAAVTEHFLPFLNGKRFYFGVKNKHEADIRGFGDWPFLVDPFLRLWLGANA 379
Qy 314 FRTSHYPYSEEWLADRLGILVIDEAPHVGI--TRYHNPETQ---KIAEDNIRMTDR 368
Db 380 FRTSHYPYAEVMOQCDRYGIVVIDESPQVGI VLVESVNSLSQHLNLMELVAV---R 435
Qy 369 HKNHPSVIMSVANPESNHPDAEGFPKALYETANEMOTRPRVNVSMMDADERTRDA 428
Db 436 DKNHPRVVMWSVANPASFLLKPRAGYFFKTLIAHTKALDPSRVEFTV---NSNYEADLG 491
Qy 429 LKYPDIVCVNRYGKRYIQGRIEBGLQALBKDIEELYARHKRPIVTFEGADALGHIYD 488
Db 492 APYVDVICVNSYSYVHDYGNHEVITQLQATQFENWYRTYQKPIIQSEYGDITJAGFHOD 551
Qy 489 PPMFSESEYQALVKTIRLLP---KDYIIGTHVMAFADFKTPQNVBRPILNHKGFTFR 545
Db 552 PPLMSESEYQGLLEQ-YHLVLDQKRYVGVSELIMNFRADPMTNOSPQVWGNKKGITFR 610
Qy 546 DRQKLVAVLR-RLW 560
Db 611 ORQPKAFLRLERYW 626
RESULT 14
BgLR_RAT ID BgLR_RAT STANDARD; PRT; 648 AA.
AC P06760;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=Guab; Synonyms=Gu; Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Preputial gland;
RX MEDLINE=87016933; PubMed=3463967;
RA Nishimura Y., Rosenfeld M.G., Kreibich G., Gubler U., Sabatini D.D., Adesnik M., Andry R.;
RT "Nucleotide sequence of rat preputial gland beta-glucuronidase cDNA and in vitro insertion of its encoded polypeptide into microsomal membranes.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:7292-7296(1986).
RN [2]
RP SEQUENCE OF 14-648 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88183378; PubMed=3355337;
RA Powell P.P., Kyle J.W., Miller R.D., Pantano J., Grubb J.H., Sly W.S.;
RT "Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons and expression of a chimeric protein in COS cells.";
RL Biochem. J. 250:547-555(1988).
CC -1- FUNCTION: Plays an important role in the degradation of dermatan and keratan sulfates.
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol + D-glucuronate.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- PTM: Undergoes a post-transcriptional proteolytic cleavage near its C-terminal end, which reduces its size by approximately 3 kDa. The site of this cleavage has as yet not been determined.
CC -1- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M13962; AAA41228.1; -;
DR EMBL; Y00717; CAA68705.1; -;
DR PIR; A25047; A25047.
DR HSSP; P08236; 1BHG.
DR RGD; 2772; Guab.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_219.
DR InterPro; IPR006104; Glyco_hydro_258.
DR InterPro; IPR006103; Glyco_hydro_2T1M.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRILASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KM Glycoprotein; Glycosidase; Hydrolase; Lysozyme; Signal.
FT SIGNAL; 1 22
FT CHAIN; 23 648 Beta-glucuronidase.
FT ACT SITE; 447 447 Proton donor (By similarity).
FT CARBOHYD; 172 172 N-linked (GLCNAc...) (Potential).
FT CARBOHYD; 416 416 N-linked (GLCNAc...) (Potential).
FT CARBOHYD; 591 591 N-linked (GLCNAc...) (Potential).
FT CARBOHYD; 627 627 N-linked (GLCNAc...) (Potential).
FT CONFLICT; 14 14 Q -> E (in Ref. 2).
FT CONFLICT; 21 21 V -> L (in Ref. 2).
FT CONFLICT; 487 487 M -> L (in Ref. 2).
SQ SEQUENCE 648 AA; 74793 MW; 5AD8E85234F0907E CRC64;
Query Match 29.8%; Score 894.5; DB 1; Length 648;
Best Local Similarity 36.1%; Pred. No. 1e-50;
Matches 225; Conservative 98; Mismatches 210; Indels 91; Gaps 20;
Qy 1 MVRQGRNKRKIRILLNGVWN-----LEYTSKORPAVAGSNME- 38
Db 27 MLFPEKTPSRRLKVLQDLGMSFRADYSNNRLQGFQKQWYROPRLRSGLTIDMPVBSFNDI 86
Qy 39 -QYODLCYEEGPFYKTFYVVKXLSQKIRLRYPAAVNTD---CEVFLNGEKVGENHIEY 94
Db 87 TQELALNFIQWYWERAVLPQKWTQDTRRVLRINSATYAVVWVNGIHVHEBQGH 146
Qy 95 LPPEVDVTKVYKSG--ENELRVVENRLKVGFSPSKVP-----DSGTHVGFPGSF 143
Db 147 LPFADIDIKVLQSGPLTFTRVTAIINNTL-----PYLTPGTIVYKTDPSMYPKGYF--V 200
Qy 144 PRANPDRFPYGGIIRPVLIETFDHARTIDIVNDTSSESPREKKLKVKKIVSEBAVQGE 203
Db 201 QDISFDEFNVAGLRSVVLYTPTTPTTIDITVTT---DVDKDVGLVNWYISV-QGSDHFQ 256
Qy 204 MTKLGESEKIRTSNRFVEGEFTLENNARFW---SLEDP-YLYPLKYLE-----ELEKDEY 253
Db 257 LEVLLDBDQKIVARGCNBQGLAVPRNAHLMWRYLMHBPALVLSLEVTALQTAAGSVDFY 316
Qy 254 TLDIGIRTSMDKRLYLNGKRVPLKGFQKHEEPVLQGTFFYPLMIDPNLLKMINNS 313
Db 317 TLPVGIRTVAAVTEHFLPFLNGKRFYFGVKNKHEADIRGFGDWPFLVDPFLRLWLGANA 376
Qy 314 FRTSHYPYSEEWLADRLGILVIDEAPHVGI--TRYHNPETQ---KIAEDNIRMTDR 363
Db 377 FRTSHYPYSEEWLQCDRYGIVVIDESPQVGI VLVESVNSLSQHLNLMELVAV---R 431
Qy 364 RMIRHKNHPSVIMSVANPESNHPDAEGFPKALYETANEMOTRPRVNVSMMDADERTRDA 423
Db 432 ---RDKNHPVVMWSVANPASFLLKPRAGYFFKTLIAHTKALDPSRVEFTV-----N 480

QY 424 TR---DVALKFEDIVCVNRYGYWYIYOGRIEGLQALEKDIIEELVYARRKPIFVTEBGAD 480
 DB 481 TRYADMGAPYVDIVCVNSYISWYHDYGHLEVIQLTSCQENNYKMYQKFIQISEGAD 540
 QY 481 AIAGIHVDPMPQFSESEYQALVEKTIIRLL---KKDYIIGHVAFADFKTPQVRRPIL 537
 DB 541 AVSGIHEDPMPFSESEYQALVE---NYHLIDKKEKVEYIGELINWADFMVNOPLAVTG 599
 QY 538 NHKVFTRDQPKLVAVLR-RLW 560
 DB 600 NKKGIIFTRQRPKMAAFILRERYW 623

RESULT 15
 BGLR_CANFA
 ID_BGLR_CANFA STANDARD; PRT; 651 AA.
 AC 018835;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Beta-glucuronidase precursor (EC 3.2.1.31).
 GN Name=GUSB;
 OS Canis familiaris (dog).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N. A. AND VARIANT MPS VII HIS-166.
 RA MEDLINE=98190525; PubMed=9521879;
 RA Ray J., Bouvet A., Desanto C., Fyfe J.C., Xu D., Wolfe J.H.,
 RA Aguirre G.D., Patterson D.F., Haskins M.E., Henthorn P.S.;
 RT "Cloning of the canine beta-glucuronidase cDNA, mutation
 RT identification in canine MPS VII, and retroviral vector-mediated
 RT correction of MPS VII cells";
 RL Genomics 48:248-253(1998).
 CC -1- FUNCTION: Plays an important role in the degradation of dermatan
 CC and keratan sulfates (By similarity).
 CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +
 CC D-glucuronate.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis
 CC type VII (MPS VII), an inherited disease reported in humans, mice,
 CC cats, and dogs.
 CC -1- SIMILARITY: Belongs to family 2 of glycosyl hydrolases.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

CC -----
 DR EMBL; AF019759; AAC48809.1; -.
 DR HSSP; P08236; 1BHG.
 DR InterPro; IPR008979; Gal bind like.
 DR InterPro; IPR006101; Glyco_hydro_2.
 DR InterPro; IPR006102; Glyco_hydro_21g.
 DR InterPro; IPR006104; Glyco_hydro_25B.
 DR InterPro; IPR006103; Glyco_hydro_21TM.
 DR Pfam; PF02703; Glyco_hydro_2; 1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDLASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
 KW Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysosome;
 KW Signal.
 FT SIGNAL 1 22 By similarity.
 FT CHAIN 23 651 Beta-glucuronidase.
 FT ACT_SITE 450 450 Proton donor (By similarity).
 FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).

FT CARBOHYD 419 419 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 630 630 N-linked (GlcNAc...) (Potential).
 FT VARIANT 166 166 R -> H (in MPS VII; loss of activity).
 SQ SEQUENCE 651 AA; 74433 MW; E8991B1E65C60120 CRC64;
 Query Match 29.6%; Score 888; DB 1; Length 651;
 Best local Similarity 35.7%; Pred. No. 2, 8e-50;
 Matches 220; Conservative 98; Mismatches 226; Indels 72; Gaps 18;
 QY 1 MVRPQNRKRFIIILNGVNLVETSKD-----RPIAVGSWNEQ 39
 DB 27 MLYRSPSPSRERKDLGLWSFRADFGRQGFEGQWYAPRLAESGPTLDMVPSSFNDV 86
 QY 40 YOD--LCYERGPFTYKTFVPPKLSO---KHRLFAAVNDCEFLNKEKGEHIEY 94
 DB 87 GQDRQLRSFVGWVWYEREATLPRWSQDPTGRVLLIGSNHYVAIVWVGVAHEHGGH 146
 QY 95 LPPEVDVTGKVKSG--ENELRVVENRLKVGSPSKVP-----DSGTHVGFPGSF 143
 DB 147 LPPEADISKLVGSGPLSSCRITLAINNTLT---PHTLPRTIVYKTDASKYKGF--V 200
 QY 144 PRANFPFPYGGIIRPVLIETFDHARILDIWDTSESEPEKKLGKVKKIEVSEAVQGE 203
 DB 201 QNTYFPFPFYAGIHRFVLLYPTFTYIDITVTVGVQD-DTGLVDVQIFVQSGSEHF---Q 256
 QY 204 MTKLGESEKIRTSRPFGEFLENAFW-----SLSDP-VLYPLKVELE-----K 250
 DB 257 LEVYLDEBERKVAQGTSGQRLQVPNVHLMWYLNHEHRAVYLSLEVRITLQMAAGPV 316
 QY 251 DEYTLDIGRITISWDEKRLYLNGKPVFLKGFGEHEFPVLGQSTFPYPLMIKDFNLKWIN 310
 DB 317 DFYTLFVGIRTVAVTERQFLNGKFPYFGVNNHEBADIIRGKGFDMPLVKDFNLRLWIG 376
 QY 311 ANSFRTSHYPSSEMLDLADRGLIYIDAPHYGI--TRYHNPETQKIAEDNIRKMDR 368
 DB 377 ANAFRTSHYPAEYVQLDRIYGIVIDSPGIMLVOSYSNVSQHNILE-TWGEVLVR 435
 QY 369 HKHPSVIVMSVANEPSNHPDAEGFKALYETANEMDRTPVWVSMVDADERTRDVA 428
 DB 436 DKHPSVIVMSVANEPTSLKPAAYFKTLIAHTKALDSRPVTPVT---NSNYEADLG 491
 QY 429 LKXFDIVCVNRYGYWYIYOGRIEGLQALEKDIIEELVYARRKPIFVTEBGADAIAGIHYD 488
 DB 492 APYVDIVCVNSYISWYHDYGHLEVIQLQATEPENNYRTYQKRIQSEGAETIAGFHQD 541
 QY 489 PPMSESEYQALVEKTIIRLL---KKDYIIGHVAFADFKTPQVRRPILNHKVFTR 545
 DB 552 PPMSESEYQKGLLEQ-YHLVLQKKEKVEYIGELINWADFMVNOPLAVTG 610
 QY 546 DRQPKLVAVLR-RLW 560
 DB 611 QROPKMAAFILRERYW 626

Search completed: January 24, 2005, 07:52:35
 Job time : 109 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 24, 2005, 01:17:00 ; Search time 7288 Seconds
(without alignments)
10959.427 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689
Sequence: 1 atggttaagaccgcaacgaacaa.....gaagactgtgagatgagatc 1689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 2364489745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Genembl: *
1: gb_ba: *
2: gb_bcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sbs: *
12: gb_sy: *
13: gb_un: *
14: gb_vt: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	6	AR428149 Sequence
2	1688	99.9	12583	1	AE001766 Sequence
3	194	11.5	2153	4	AF012423 Felis cat
4	192.4	11.4	2014	4	AF012424 Felis cat
5	190.2	11.3	2199	4	AF019759 Canis fam
6	174.4	10.3	2155	9	AF084552 Chlorocob
7	174	10.3	2128	6	COS75703 Sequence
8	174	10.3	2182	9	BC014142 Homo sapi
9	174	10.3	2321	3	BT010073 Drosophi1
10	172.4	10.2	1956	6	AR344264 Sequence
11	172.4	10.2	1956	6	AX147652 Sequence
12	172.4	10.2	2191	9	HUMGLCB
13	166.8	9.9	2391	10	RNGLCB
14	166.8	9.9	2472	6	AX827491 Rat mRNA fo
15	166.8	9.9	2472	10	RATGLCB
16	159.6	9.4	1947	6	AR344265 Sequence
17	159.6	9.4	1947	6	AX147654 Sequence
18	159.6	9.4	2456	10	MUSGUS
19	159.6	9.4	2456	10	MUSGUS

20	159.6	9.4	2456	10	MUSGUS	M63836 Mouse beta-
21	159.6	9.4	2492	10	BC071226	BC071226 Mus muscu
22	157.8	9.3	2454	10	MUSGLCB	M19279 Mouse beta-
23	156.4	9.3	2455	10	MUSGLC	J03047 Mouse beta-
24	142.2	8.4	23078	6	CQ363720	CQ363720 Sequence
25	142.2	8.4	60265	1	AE017283	Continuation (26 o
26	141.4	8.4	1809	12	AF354047	AF354047 Synthetic
27	141.4	8.4	1887	6	AR210802	AR210802 Sequence
28	141.4	8.4	1887	6	BD074927	BD074927 Microbia
29	141.4	8.4	7018	12	AY452736	AY452736 Reporter
30	141.4	8.4	11846	12	AF354045	AF354045 Binary ve
31	141.4	8.4	11921	12	AF354046	AF354046 Binary ve
32	141.2	8.4	2235	11	BV175016	BV175016 eqm76123
33	139	8.2	10775	1	AE006894	AE006894 Sulfolobu
34	138.6	8.2	2352	3	BT012475	BT012475 Drosophi1
35	138.6	8.2	2445	3	AY069329	AY069329 Drosophi1
36	130.4	7.7	1888	6	AR428153	AR428153 Sequence
37	127.6	7.6	93590	3	AC084452	AC084452 Caenorhab
38	123.6	7.3	301450	1	AP003185	AP003185 Clostridi
39	122	7.2	3448	1	CPE420784	AJ420784 Clostridi
40	115.6	6.8	2105	1	AF305917	AF305917 Escherich
41	115.6	6.8	300330	1	AP005222	AP005222 Corynebac
42	114	6.7	1814	6	BD185061	BD185061 Nucleic a
43	114	6.7	2103	1	AF305918	AF305918 Escherich
44	114	6.7	10679	1	AE005385	AE005385 Escherich
45	114	6.7	257071	1	AP002558	AP002558 Escherich

ALIGNMENTS

RESULT 1	AR428149	1689 bp	DNA	linear	PAT 18-DEC-2003
LOCUS	AR428149	Sequence 14 from patent US 6641996.			
DEFINITION	AR428149				
ACCESSION	AR428149.1	GI:40187542			
VERSION					
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1689)				
AUTHORS	Jefferson R.A. and Mayer, J.E.				
TITLE	Microbial .beta.-glucuronidase genes, gene products and uses				
JOURNAL	Patent: US 6641996-A 14 04-NOV-2003;				
FEATURES	Location/Qualifiers				
source	1..1689				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	99.9%; Score 1688; DB 6; Length 1689;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 1689;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	ATGTTAAGACCGCAACGAACAGAGATTATTTCTTATCTTGAATGAGATTGGAAT	60		
DB	1	ATGTTAAGACCGCAACGAACAGAGATTATTTCTTATCTTGAATGAGATTGGAAT	60		
QY	61	CTTGAAGTAAACGAGAAAGACAGACCAATCCCGTTCTGGAAGCTGGAATGAGAGATAC	120		
DB	61	CTTGAAGTAAACGAGAAAGACAGACCAATCCCGTTCTGGAAGCTGGAATGAGAGATAC	120		
QY	121	CAGATCTGTGCTCAAGAGAGACCCCTTCACTTCAAAAACACCTTCTAGCTTCCGAAG	180		
DB	121	CAGATCTGTGCTCAAGAGAGACCCCTTCACTTCAAAAACACCTTCTAGCTTCCGAAG	180		
QY	181	NAACCTTCAAAAACACATCAGACTTACTTGTGCTGGTGAACACGAGCTCGAGGATC	240		
DB	181	NAACCTTCAAAAACACATCAGACTTACTTGTGCTGGTGAACACGAGCTCGAGGATC	240		
QY	241	TTCCTCAACGAGAAAGTGGAGAGATACATCATTAATACCTTCCCTTGGAGATGAT	300		

Db	241	TTCTCCAAACGAGAGAAAGATGGGAGAAATCACTTGAATACCTTCCTTCGAACTAGAT	300
Oy	301	GTGACGGGGAAAGTGAATCCGGAGAGAACGACTCAGGGTGTGTTGAGAACGATTG	360
Db	301	GTGACGGGGAAAGTGAATCCGGAGAGAACGAACTCAGGGTGTGTTGAGAACGATTG	360
Oy	361	AAAGTGGGAGATTTCCCTCGAAGGTTCCAGACACCGGCACTCACCGTGGGATTTTTT	420
Db	361	AAAGTGGGAGATTTCCCTCGAAGGTTCCAGACACCGGCACTCACCGTGGGATTTTTT	420
Oy	421	GGAAATTTTCCACTCTGCAACTTCGACTTTCCTCCCTACGGTGAATCAAAAGCCTGT	480
Db	421	GGAAATTTTCCACTCTGCAACTTCGACTTTCCTCCCTACGGTGAATCAAAAGCCTGT	480
Oy	481	CTGATAGAGTTTCAACAGACACGCGAGGATCTCGACATCTGGGTGAGACACGATGACT	540
Db	481	CTGATAGAGTTTCAACAGACACGCGAGGATCTCGACATCTGGGTGAGACACGATGACT	540
Oy	541	GAAACCGAGAGAAACCTTGAAANAAGTGAAGATGAAAGTCTCAGAAAGACGGTG	600
Db	541	GAAACCGAGAGAAACCTTGAAANAAGTGAAGATGAAAGTCTCAGAAAGACGGTG	600
Oy	601	GGAACGAGAGATGACGATCAACTTGGAGAGGAAAGAAAGATTGAAACATCCACAGA	660
Db	601	GGAACGAGAGATGACGATCAACTTGGAGAGGAAAGAAAGATTGAAACATCCACAGA	660
Oy	661	TTTCGTGCAAGGGGAGTTACTCTCGAAAACGCGACAGTTCTGGAGGCTCGAAGATCATAT	720
Db	661	TTTCGTGCAAGGGGAGTTACTCTCGAAAACGCGACAGTTCTGGAGGCTCGAAGATCATAT	720
Oy	721	CTTTATCTCTCTCAGGTGGAACCTTGAAAAGAAGGATCACTCTGGAATCGGAATCAGA	780
Db	721	CTTTATCTCTCTCAGGTGGAACCTTGAAAAGAAGGATCACTCTGGAATCGGAATCAGA	780
Oy	781	ACGATCAGCTGGGAGAGAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGAC	840
Db	781	ACGATCAGCTGGGAGAGAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGAC	840
Oy	841	TTTGGAAAGCAGAGAAATCCCGGTTTGAGGAGAGGCACTTTTATCATTTGATGATA	900
Db	841	TTTGGAAAGCAGAGAAATCCCGGTTTGAGGAGAGGCACTTTTATCATTTGATGATA	900
Oy	901	AAAGCTTCAACCTTGTGAAGTGAATCAACGGGAATTTCTTCAGGACCTTCACATATCT	960
Db	901	AAAGCTTCAACCTTGTGAAGTGAATCAACGGGAATTTCTTCAGGACCTTCACATATCT	960
Oy	961	TACAGTGAAGGTGCTGATCTTCGCGACAGACTCGGAATCTTGTGATGACGAAGCC	1020
Db	961	TACAGTGAAGGTGCTGATCTTCGCGACAGACTCGGAATCTTGTGATGACGAAGCC	1020
Oy	1021	CCGCACTTTGGTATACAAAGATCACTACATATCCCGAGACTCAGAAAGATGACAGAAAC	1080
Db	1021	CCGCACTTTGGTATACAAAGATCACTACATATCCCGAGACTCAGAAAGATGACAGAAAC	1080
Oy	1081	AACATAAAGAAATATGATGACAGACAGACAAACATCTCCAGTGTATCATGTGAGGTG	1140
Db	1081	AACATAAAGAAATATGATGACAGACAGACAAACATCTCCAGTGTATCATGTGAGGTG	1140
Oy	1141	GCGAAGCAACGAGTCCAAACATCCAGACGCGAGGGGTTTCTTCAAAAGCCTTTATGAG	1200
Db	1141	GCGAAGCAACGAGTCCAAACATCCAGACGCGAGGGGTTTCTTCAAAAGCCTTTATGAG	1200
Oy	1201	ACTGCGAATGAAATGATCGAACAGCCCGTTGTCATGATGATGATGATGAGACGACCA	1260
Db	1201	ACTGCGAATGAAATGATCGAACAGCCCGTTGTCATGATGATGATGATGAGACGACCA	1260
Oy	1261	GACGAGAGAACAGAGACGTGGCGCTGAAAGTACTTCGACATCGTCTGTGTGAACAGTAC	1320
Db	1261	GACGAGAGAACAGAGACGTGGCGCTGAAAGTACTTCGACATCGTCTGTGTGAACAGTAC	1320
Oy	1321	TACGGCTGTACTTATCAGGAAAGATGAAAGGACTTCMAAGCTCTGAAAAAGAC	1380
Db	1321	TACGGCTGTACTTATCAGGAAAGATGAAAGGACTTCMAAGCTCTGAAAAAGAC	1380

Db	1321	TAGCGCTGTGATACATCTCATCGGGAAAGGATGAAGAAGAGACTTCGAAGCTTCGAAAAAAGAC	1380
Oy	1381	ATAAAGAAGCTCTTATGCAGAAGCACAGMAAACCCATTCTTTGTCAACAGAAATTCGGTGCGGAC	1440
Db	1381	ATAAAGAAGCTCTTATGTGAAGGACACAGAAACCCTCTTTGTCAACAGAAATTCGGTGCGGAC	1440
Oy	1441	GCGATAGCTGGCATCCACTACGATCCAGTCCACTCMAATGTCTTCGCCAAGAGTACCAAGACAGAG	1500
Db	1441	GCGATAGCTGGCATCCACTACGATCCAGTCCACTCMAATGTCTTCGCCAAGAGTACCAAGACAGAG	1500
Oy	1501	CTCGTTGAAAAGACGATCAGGCTCTCTTTGAAAAAGACATACATCATCGGAACACAGCTG	1560
Db	1501	CTCGTTGAAAAGACGATCAGGCTCTCTTTGAAAAAGACATACATCATCGGAACACAGCTG	1560
Oy	1561	TGGGCTTTTGACAATTTTAAGACTCCTCAGAAATGTGAGAAAGACCATTCTCAACCAAGAAG	1620
Db	1561	TGGGCTTTTGACAATTTTAAGACTCCTCAGAAATGTGAGAAAGACCATTCTCAACCAAGAAG	1620
Oy	1621	GGTGTTCACACAGAAGACAGACCAACCCAACTCGTTGCTCATGTACTGAGAAGACTGTGG	1680
Db	1621	GGTGTTCACACAGAAGACAGACCAACCCAACTCGTTGCTCATGTACTGAGAAGACTGTGG	1680
Oy	1681	AGTAGAGTT	1689
Db	1681	AGTAGAGTT	1689
RESULT 2	AE001766/c		
LOCUS	AE001766	12583 bp	DNA linear BCT 04-JUN-2004
DEFINITION	Thermotoga maritima MSB8 section 78 of 136 of the complete genome.		
ACCESSION	AE001766	AB000512	
VERSION	AE001766.1	GI:9981600	
KEYWORDS			
SOURCE ORGANISM	Thermotoga maritima MSB8 Thermotoga maritima MSB8 Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.		
REFERENCE AUTHORS	1 (bases 1 to 12583) Nelson,K.E., Clayton,R.A., Gill,S.R., Gwin,M.L., Dodson,R.J., Hatt,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Uteerback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,P., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., Eisen,J.A., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.W. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima Nature 399 (6734), 323-329 (1999)		
TITLE	2 (bases 1 to 12583)		
JOURNAL MEDLINE	Nature 399 (6734), 323-329 (1999)		
PUBMED	99287316		
REFERENCE	10360571		
AUTHORS	2 (bases 1 to 12583) Nelson,K.E., Clayton,R.A., Gill,S.R., Gwin,M.L., Dodson,R.J., Hatt,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A., McDonald,L., Uteerback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M., Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A., Richardson,P., Heidelberg,J., Sutton,G.G., Fleischmann,R.D., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.W. Direct Submission Submitted (01-JUN-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA Location/Qualifiers 1. 12583 /organism="Thermotoga maritima MSB8" /mol_type="genomic DNA" /strain="MSB8" /db_xref="taxon:243274" 81..1607 /locus_tag="TM1058" 81..1607 /locus_tag="TM1058" /note="similar to GH:AE000666 percent identity: 48.70; identified by sequence similarity; putative /codon_start=-1		
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			

```

/transl_table=11
/product="glutamate synthase-related protein"
/protein_id="AAD36135.1"
/db_xref="GI:4981601"
/translation="MSGLCACVCVCGCTGCCTWLSFGRREVLYPGPGDITAGAVKQ
YPTVSHNLTIGYVARGAGLPGVEVGPDTAIFTVDTTTEGMDIKMKVPIETGA
LGSIRAEKMMNDHILAVGAALISGITVVCGEVAVGDPDLELNSGVKKSPGLDRRIE
YKRYHDEYGEYILIQMNVEDTRLCVAVYINKHGTEELKMGQAKSIGEIKYRSL
ERALBLKRGITVLDPPELPVQRAFKGEETKEFRHSRLSPVSESLKEVERKRL
GPKRITLKTGAVSAVELMALRYGAEARVDLITVDAGCGMSPPMNMWGIPTFY
LEALYQPAEKLISRGISIPDIALINGFSTEDVLKAIAMGSPYKAVACMRALMIPIA
VQKNIGEMIKSGNLPTKVSXKGTVEEIPVTEBELRSRFGSEEVKPLPLGAGIYTF
VQKFTGLQQLMAGARKFRLSALSKDILALTKAAELISGIPYMWESYRDAERILSE
"
complement(1650..1913)
/locus_tag="TM1059"
complement(1650..1913)
/locus_tag="TM1059"
/notes="similar to SP:P45693 PID:862985 GB:AL009126 percent
identity: 48.70; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="spovs-related protein"
/protein_id="AAD36136.1"
/db_xref="GI:4981602"
/translation="WEVLKVSXSQSDPNKVAGALGVREHGAETIQALGAVNQAVK
AIALRGYGLASGIDLVFPVPAFTDVEIENKRTAIKFLVFPKS"
complement(1962..3272)
/locus_tag="TM1060"
complement(1962..3272)
/locus_tag="TM1060"
/notes="similar to GB:AE000782 percent identity: 49.74;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36144.1"
/db_xref="GI:4981610"
/translation="WNISIEGALAVLINOFGGAYLTGYFLMNGASSPFIQLGSPISIP
FLANTLQLLTFSFSRLKSRKQIIVPLMWTARTSILLPVPAIKGILLAYLVFYI
QIAGLSAPLQSNMWSLDPKMDIGSYGFGRNLHGAVQIDPAMFLAGAILDSLENKK
FGRTLPLAGSLGANGVELKIOYRPPKPREASVTAKVFKLLKEHFRNPLGPA
FNRPAIGCTYTYINMLKEVPFSLQISVANAGMFIGTLPOPFMGLGRYGFQYR
LKVCLMIAIYILMTLTPRSFLYVFIQIITIGLPTVAGTSLQVFTLMTYAPSSLKT
BAFVFNLSLMSLFASSLVASVVASLNSLDPGISAIRLMTFISPLASAAAYII
SRMDLGTQKDSLQIAVKESFFSGTVFWIRERLNTLNIFFRKR"
complement(3382..4545)
/locus_tag="TM1061"
complement(3382..4545)
/locus_tag="TM1061"
/notes="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36137.1"
/db_xref="GI:4981603"
/translation="NRSWTDVRYWISLRLKICEMPELSCASDLKSKSPVEGKSESE
RRKFTLELRIGFISPLPELNRSPSTDPEERKIATRLSEPAVKSIDVATPNTK
DYNMFKRGROPLVDAAPLVEALILAPKYMDDLVSTKRLIRELKAIRKIEPFSNK
LIFSAMITFFPFGSEMDSTVDLILNKVESMKGDGAYDGGPEFRDYNSPVITP
MTIOLVRLISEKTEWKLKYVAVLRARAVVVERMSIPGTPPIIGRSITRYTAVF
HTLQSLHLHLPASLSPAQVCALTAVALRIFFENPSIFDENGMLKIGISQPSLGE
EYITGLSLYCTVFLPLGLPTSDPFMDPKCKTKNKVMEGEDVADPRALED"
complement(4542..6233)
/locus_tag="TM1062"
complement(4542..6233)
/locus_tag="TM1062"
/notes="similar to PID:642973 PID:14446 PID:412358 percent
identity: 57.73; identified by sequence similarity;
putative"
/codon_start=1

```

```

/transl_table=11
/product="beta-glucuronidase"
/protein_id="AAD36143.1"
/db_xref="GI:4981609"
/translation="MVRPQRNKKRFFILILNGWNLAEVTSKDRPIAVPGSNBOYODLC
YEEBPFYKTFYFPEKELSOGLILYPAAVTDEVFNGSKVESENHIEYLFPEVDY
GKVSXGEMLRVVENNLTKVGGPSKVDGOSTHIVGFGSPSPANPDFEPVGGIIRPV
LIETDARILIDIVDTSSEPEKKLQKVKIKIVSEBAVQEMTITKIGEEKIIRTS
NRVEGEFTILSNAPWSLEDYLYPLKVELKDEYTLIDIGRTISWDEKRLYLNGKPV
FLKPFGEKEEPVVLQGFPEYPLMKDNLMLMINANSRTHSYSPVSEMDLADRLGI
LVIDEAPHVIGITRTHYNPEYQKIMEDNIRMIIDHKHPSVIMSVANEPSPNDAR
GEFPALETANEMRTRPVVWVMSMDAPDEBTTRVALKAPDVCVNRYYGQIRI
BEGIOALEKOTIEELVARRKPIFTLEGADLIDGIVHDPMPSEEVQAEIVERTIRL
LKKDYIITGVHAFADFKTPQNVRRPLNKHGVFTTRDROKPLVAHVLRMLSEV"
complement(6284..7309)
/locus_tag="TM1063"
complement(6284..7309)
/locus_tag="TM1063"
/notes="similar to SP:P42065 PID:677944 GB:AL009126 percent
identity: 75.31; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36138.1"
/db_xref="GI:4981604"
/translation="MALLEVNKLKKYPIIKRGRKRVGYLKAVDGISFPIEBGETL
GLVGEISGCGRTTAKLIMRGIEPTEGEIVLMDGEKVDITGLSEKELSEKGVRFLOM
IFODPVASINPRMTVRDIIABPIVNGVYKGEKEIDIVSDLRAVGLRPRYMOYRPH
AFSGGQRIAIANAIALKPKLVYCDPEPTSLDPSVQAQIITNLKDLOKENLYLPLT
SHDGVGVHITNRVAVMTVGKIVELAEIHELSSSPKRTYALSAVFPKPKRRKK
AQLGVVDPENSPSGCYFHRCPYAAIKCEYVPDLIRNVTGIDQNPHLVACHPADSLK
LEGVGM"
complement(7296..8294)
/locus_tag="TM1064"
complement(7296..8294)
/locus_tag="TM1064"
/notes="similar to SP:P42064 PID:677943 GB:AL009126 percent
identity: 72.95; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36139.1"
/db_xref="GI:4981605"
/translation="MSTLLQIKNLRTYFPFDGUVKAVDGVSPFIEBGRTLGVGSEG
CGKSVTARSITIKLSTAGRIYSGEILYNMGQMDLYKFSKEEIRKRGRIAMFQF
PMAAFSPVYITIGDQITGEMTYHFGITQOAEKRAVELLRGVGIXPEKMDISYPEYS
GMMQRAMIAVALSCNRLIADPEPTLADVTIOAVDLKDLOQEKYKAIMITN
MGVVAEMADHVWVYVLRVVASAVEELFVYPKPYSLILRSIPVVGKVRERLEVE
GDVDDPRNMPKGCSEHPRCPYMKMGICDEREPVEVEGPEHRVSCFLYGEKDGAS"
complement(8306..9409)
/locus_tag="TM1065"
complement(8306..9409)
/locus_tag="TM1065"
/notes="similar to GB:AE000657 percent identity: 67.80;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36140.1"
/db_xref="GI:4981606"
/translation="MKKREQMEKEFYVYASQQLIMRFRKKLAVIGGVLLITVFA
IFCEPAPVDONKNRYRYPAPORIHFFHFGKIGVYGVYTVVLETRIRIYKED
KSKIKFKPRGDEYKEMGKWKNIHIFIGVQDMSHMLTGDSIGRMLSIITGAPR
STSIGLGVLSPLILGVALIGISGIFGGAQDNFQRTIETIKSIPTPPLIALSAAP
QNMPLKRVFVIVILISLTGWTDLARVRSRPSLRBEDFVMAKFEWASARLIIFRH
MLPSFMGHLIASITLSIPGMIGLSTSLGLGLRPPVISWGLLOBOAONLTVALVP
WLLIPVVFVITTVICFNFGVGLDADDPYANM"
complement(9420..10403)
/locus_tag="TM1066"

```

gene

CDS complement(9420...10403)

/locus_tag="TM1066"
/note="similar to SP:P42062 PID:677946 GB:AL009126 percent
identity: 62.93; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36141.1"
/db_xref="GI:4981607"

Query Match 99.9%; Score 1688; DB 1; Length 12583;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGTGAAGACCGCAACGAAACAGAGATTATTTCTTATCTTGAATGAGTTGGAAT 60
6233 ATGTGAAGACCGCAACGAAACAGAGATTATTTCTTATCTTGAATGAGTTGGAAT 6174
61 CTGAAGTAAACGACAAAGACAGACCAATCGCGTTCTCGAAGCTGGAATGAGCACTAC 120
6173 CTGAAGTAAACGACAAAGACAGACCAATCGCGTTCTCGAAGCTGGAATGAGCACTAC 6114
121 CAGGATCTGTGCTACGAAAGAGACCTTCACTACAAACCACTTCTAGTCCGAAG 180
6113 CAGGATCTGTGCTACGAAAGAGACCTTCACTACAAACCACTTCTAGTCCGAAG 6054
181 NAACCTTCACAAAAACATCAGACTTACTTCTGCGGTGAACACGAGCTGCGAGTTC 240
6053 NAACCTTCACAAAAACATCAGACTTACTTCTGCGGTGAACACGAGCTGCGAGTTC 5994
241 TTCTCAACGGAAGAAAGTGGAGAGATCACTTAATACCTTCCCTCGAAGTGAAT 300
5993 TTCTCAACGGAAGAAAGTGGAGAGATCACTTAATACCTTCCCTCGAAGTGAAT 5934
301 GTGACGGGAAAGTGAATCCGAGAGACGAACTCAGGGGTGTTGTTGAACAGATTG 360
5933 GTGACGGGAAAGTGAATCCGAGAGACGAACTCAGGGGTGTTGTTGAACAGATTG 5874
361 AAAGTGGAGAGATTTCCCTCGAAGATTCAGACGCGCACTCACACCGTGGAGTTT 420
5873 AAAGTGGAGAGATTTCCCTCGAAGATTCAGACGCGCACTCACACCGTGGAGTTT 5814
421 GGAAGTTTCCACCTGCAAACTTGCATCTTCCCTACGAGTGAATCATAGGCGCTGT 480
5813 GGAAGTTTCCACCTGCAAACTTGCATCTTCCCTACGAGTGAATCATAGGCGCTGT 5754
481 CTGATAGATTACAGACCAACGAGAGATCTCGACATCTGGGTGACACAGATGACT 540
5753 CTGATAGATTACAGACCAACGAGAGATCTCGACATCTGGGTGACACAGATGACT 5694
541 GAACCGGAAAGAACTTGGAAAAAGTGAAGTGAAGTGAAGTCTCAGAAAGCGGTG 600
5693 GAACCGGAAAGAACTTGGAAAAAGTGAAGTGAAGTGAAGTCTCAGAAAGCGGTG 5634
601 GGAAGAGAGATGAGATCAAACTTGGAGAGAGAGAAAAAGATTAGACATCCACAGA 660
5633 GGAAGAGAGATGAGATCAAACTTGGAGAGAGAGAAAAAGATTAGACATCCACAGA 5574
661 TTGCTGAAGGGGAGTTCTATCTTGAAGACGCGAGTTCTGAGCCTCGAAGATCCAT 720
5573 TTGCTGAAGGGGAGTTCTATCTTGAAGACGCGAGTTCTGAGCCTCGAAGATCCAT 5514
721 CTTTATCTCTCAAGGTGGAACCTTGAAGAGAGATCACTCTGAGCATCGGAATCAGA 780
5513 CTTTATCTCTCAAGGTGGAACCTTGAAGAGAGATCACTCTGAGCATCGGAATCAGA 5454
781 ACATCAGCTGGGACGAGAGAGGCTCTATCTGAAACGGGAAACCTGCTTTTGAAGGGC 840
5453 ACATCAGCTGGGACGAGAGAGGCTCTATCTGAAACGGGAAACCTGCTTTTGAAGGGC 5394
841 TTTGGAAGACGAGAGATTTCCCGTTCTGGGGCAGGCACTTTATCCATGATGATA 900

5393 TTTGGAAGACGAGAGATTTCCCGTTCTGGGGCAGGCACTTTATCCATGATGATA 5334
901 AAAGACTTCAACCTTGTGAAGTGAATCAACGCGAATTTCTTTCAGACCTTCACTATCT 960
5333 AAAGACTTCAACCTTGTGAAGTGAATCAACGCGAATTTCTTTCAGACCTTCACTATCT 5274
961 TACAGTGAAGAGTGGCTGAGTCTTGGCCAGACAGACTCGGAATCTTGTGATAGCAAGCC 1020
5273 TACAGTGAAGAGTGGCTGAGTCTTGGCCAGACAGACTCGGAATCTTGTGATAGCAAGCC 5214
1021 CCGCAGTTGGTATCAAGATGACACTCAATCCCGAGCTCAGAAAGATAGCAAGAGAC 1080
5213 CCGCAGTTGGTATCAAGATGACACTCAATCCCGAGCTCAGAAAGATAGCAAGAGAC 5154
1081 AACATGAAGAGATGATGACAGACAAAGAACATCCAGTGTGATCATGTGAGTGTG 1140
5153 AACATGAAGAGATGATGACAGACAAAGAACATCCAGTGTGATCATGTGAGTGTG 5094
1141 GCGAAGAACCGAGTCCAAACCATCCAGACGGGAGGTTTCTTCAAAAGCCCTTATGAG 1200
5093 GCGAAGAACCGAGTCCAAACCATCCAGACGGGAGGTTTCTTCAAAAGCCCTTATGAG 5034
1201 ACTGCCAATGAATGATGATGAAACGCGCCGTTGTGATGATGATGATGAGCAACA 1260
5033 ACTGCCAATGAATGATGATGAAACGCGCCGTTGTGATGATGATGATGAGCAACA 4974
1261 GACGAGAGAACAGAGACGTTGGCTGAAAGTACTTGCATGCTGTGTGAACAGGTAC 1320
4973 GACGAGAGAACAGAGACGTTGGCTGAAAGTACTTGCATGCTGTGTGAACAGGTAC 4914
1321 TACGGCTGGTATCATCATGAGGAGAGATGAAGAAAGCACTTCAAGCTCTGAAAAAGAC 1380
4913 TACGGCTGGTATCATCATGAGGAGAGATGAAGAAAGCACTTCAAGCTCTGAAAAAGAC 4854
1381 ATGAAGAGCTCTATGCAAGAGACAGAAAGCCATCTTGTGACAGAAATTCGTCGCGAC 1440
4853 ATGAAGAGCTCTATGCAAGAGACAGAAAGCCATCTTGTGACAGAAATTCGTCGCGAC 4794
1441 GCGATAGCTGGCATCTACATGATCCAAATGTTTCTCGAAGATGACAGAGAG 1500
4793 GCGATAGCTGGCATCTACATGATCCAAATGTTTCTCGAAGATGACAGAGAG 4734
1501 CTGGTGAAGAAAGCAGTACGAGGCTCTTGAAGAAAAAGATCATATATCGGAAACACAG 1560
4733 CTGGTGAAGAAAGCAGTACGAGGCTCTTGAAGAAAAAGATCATATATCGGAAACACAG 4674
1561 TGGGCTTTGAGATTTTAAGACTCTCTGAAATGTGAGAAAGCCATTTCAACCAAG 1620
4673 TGGGCTTTGAGATTTTAAGACTCTCTGAAATGTGAGAAAGCCATTTCAACCAAG 4614
1621 GGTGTTTCAAGAGACAGACCAACCCAACTGTTGCTCATGTACTGAAAGACTGTGG 1680
4613 GGTGTTTCAAGAGACAGACCAACCCAACTGTTGCTCATGTACTGAAAGACTGTGG 4554
1681 AGTGAAGTT 1689
4553 AGTGAAGTT 4545

RESULT 3
AF012423 2153 bp mRNA linear MAM 09-SEP-1999
LOCUS AF012423
DEFINITION Felis catus beta-glucuronidase (GUSB) mRNA, complete cds.
ACCESSION AF012423
VERSION AF012423.1 GI:4102550
KEYWORDS
SOURCE
ORGANISM
Felis catus (cat)
Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
1 (bases 1 to 2153)
Fyfe, J.C., Kurzhals, R.L., Laseall, M.E., Henthorn, P.S., Alur, P.R.,
Wang, P., Wolfe, J.H., Giger, U., Haskins, M.E., Patterson, D.F.,

TITLE Sun,H., Jain,S. and Yuhki,N.
Molecular basis of feline beta-glucuronidase deficiency: an animal
JOURNAL Model of mucopolysaccharidosis VII
MEDLINE Genomics 58 (2), 121-128 (1999)
PUBMED 99296826
10366443
REFERENCE 2 (bases 1 to 2153)
AUTHORS Pyfe,J.C., Kurzhals,R.L. and Henchorn,P.S.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1997) Microbiology, Michigan State University,
413 Giltner Hall, East Lansing, MI 48824, USA
FEATURES
source
1..2153
location/Qualifiers
/organism="Felis catus"
/mol_type="mRNA"
/db_xref="taxon:9685"
/tissue_type="liver"
1..2153
/gene="GUSB"
24..1979
/gene="GUSB"
/EC_number="3.2.1.31"
/note="lysosomal hydrolase; lysosomal exooligosaccharidase
in stepwise degradation of glucuronic acid-containing
oligosaccharide sidechains of proteoglycans"
/codon_start=1
/product="beta-glucuronidase"
/protein_id="AaD01498.1"
/translation="MLRGAVALMALGGLLWACGLALRGMLYPRSPSRKEKINGL
WSFPADEFENRQGEQWRYRPLRESGPTLDMPVPSFNDVGDROLRSFVGMWYE
REALPQWMODLGRVLRIGSAHYVAIWMNGVNAHEGHLPRPADSKVQSG
PLACRITAIINNTLPRLPGLTLYDTDSKYPKGFVNINDFPNVAGLRPVL
LYTPPTIYDITSTSVNODTGLVDYOI FVEGSEHQLRLLDEBGLVAQGTGR
GLQVNPANMLPMLMEHPAYLSLEVRLLAOTPAAGSDPYLLPVGIRVAATEHQ
FLINGKPEYFHVGNHEDADIRGKGDWPLKDFNLRLMAGANAFRTSHVPAEVM
QLCDRYGIVIVDESPGVIVLESYSNVLGHLLEVMELVRARKNHVAVMVAVANE
PASFLPAGVYFKTLIAATKALDPSRPPTPTNSYVADLCAPYVDVLCVNSYSYTH
DYGMETVQLQLOLATOENWRYTQPTI05SGYATGATAGFHODPPLMSEBYQGLLE
QYHLVLDQKREYVVELIMNPADFMPTN05FORVWKNKGIFTROROPKGAFLRLER
YMKLANETRYPM5AVK5QCLENSPTL"

ORIGIN

Query Match 11.5%; Score 194; DB 4; Length 2153;
Bee Local Similarity 53.1%; Fred. No. 7.3e-41;
Matches 494; Conservative 0; Mismatches 415; Indels 21; Gaps 3;

QY 751 GACGAGTACACTCTGGACATCGAATCAGAACGATCAGCTGGAGCAGAGAGGCTCTAT 810
DB 972 GACTTCTACACTCTCCCGTGGGATTCGCACGCTGCTGCACAGACACCACTTCTC 1031
QY 811 CTGAACGGGAACCTGCTCTTTTGAAGGCTTTGGAAGCAGAGAAATCCCGCTTCTG 870
DB 1032 ATCAATGGGAACCTTTCTATTCCACGGGGTCAACAGACAGAGATGACATATCCGA 1091
QY 871 GGGCAGGCGACCTTTATCATCTATGATTAAGCTTCAACTTCTGAAGTGAATCAAC 930
DB 1092 GGGAGGGGCTTGAACCTGCTGTTGAAGCACTTCAATTTGCTTCGCTGCTCGG 1151
QY 931 GCGAATTTCTCAGGACCTCTCATCTTCACTTCAAGTGAAGAGTGGCTGATCCGAC 990
DB 1152 GCCAAGCGCTCCGACCAAGTCACTACCCCTTACGGGAGAGGTATCAAGCTCTGTAC 1211
QY 991 AGACTCGGAATCTTGTATAGACGAGCCCGCAGCTTGTATCAAGATGATCA--C 1047
DB 1212 CGTATGGGATCTGTATGATGACGAGAGTCCCGGTGTGGCATGTGCTGATGAGGC 1271
QY 1048 TACAATCCCGAGACTCAGAAATAGCAGAAACAAATGAAGATATGACAGACAC 1107
DB 1272 TACAGCAACGCTCTCTGACACCACTGAGAGTATGAGAGAGCTGTGCGCAGGGAC 1331
QY 1108 AAGAACATCCCAAGTGTATGATGAGAGTGGCAGACGAACGAGTCCCAACATCA 1167

DB 1332 AAGATACACCGGCGTGTGTATGTGCTGTGGCAATGAGCCCGCTTCTCTGAAA 1391
QY 1168 GACCGGAGGGTCTTCTTCAAGCCCTTTATGACATGCAATGAATGATGAACAGCC 1227
DB 1392 CCTCGCGGTACTACTTCAAGACGCTGATTCCTCAACAAAGCCTTGATCCCTCCGG 1451
QY 1228 CCCGTTGATGAGGAGATGATGAGACGACACGACGAGAAACAGAGAGTGGCGCTG 1287
DB 1452 CCCGATCTTTGTGACCACTCCACCTATGAA-------CAGACTGGGGCG 1499
QY 1288 AAGTACTTCGACATCTGTGTGTAACAGTACTACGCGCTGTACATCTATCAGGAAAG 1347
DB 1500 CCGTATGTGACGATCATCTGTGTGATGTACTACTTGTGTATCATGATATGTGTAC 1559
QY 1348 ATGAGAGAGACCTTCAAGCTCTGGAAGAAACATAGAGCTCTATGACAGCAGCA 1407
DB 1560 ATGGAGTGATTCAGCTGACCTGCACTGGCAACCCAGTTTGAGAACTGGTATAGACCTATCAG 1619
QY 1408 AAGCCATCTTTGTCACAGATTTGGTGGACCGAGCGATGCTGGCATCTACATGATCCA 1467
DB 1620 AACCAATATATTCAGAGCGATGACGAGCAGACACCATTTGAGAGTTTCACCGAGCCA 1679
QY 1468 CCTCAATGTTCTCCGAGAGTACCAAGCAGAGCTGTTGAAAAGCAGATCAGGCTCCTT 1527
DB 1680 CCTGATGTTCAAGGAGGATATCCAGAAAGTCTGCTGAGCAGTATCATTTGTTCTG 1739
QY 1528 TTGAAA-----AAGACTACATATCGGAACACACGCTGCGGCTTTGACATTTAAG 1581
DB 1740 GATCAAAAACCAAGATATCGTGTGCGAGAGCTCATCTGAACTTGGCATTTTATG 1799
QY 1582 ACTCTGAGATGTGAGAGAGCCATTTCTCAACCAAGGCTTTTCAAGAGACAGCA 1641
DB 1800 ACTAACCGTACCGCAGAGAGTATGAGGAAATAAAAAGGAGATCTTACCGCCAGAGA 1859
QY 1642 CAACCAAACTGTTGCTCATGTACTGAGA 1671
DB 1860 CAACCAAAAGGCGACGCTTCTTTTGGA 1889

RESULT 4
AF012424
LOCUS 2014 bp mRNA linear MAM 09-SEP-1999
DEFINITION Felis catus cell-type fibroblast mutant beta-glucuronidase (GUSB)
ACCESSION mRNA, GUSB-MPS VII mutant allele, complete cds.
VERSION AF012424
KEYWORDS AF012424.1 GI:4102552
SOURCE
ORGANISM Felis catus (cat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE 1 (bases 1 to 2014)
Pyfe,J.C., Kurzhals,R.L., Lasealine,M.E., Henchorn,P.S., Alur,P.R.,
Wang,P., Wolfe,J.H., Giger,U., Haskins,M.E., Patterson,D.F.,
Sun,H., Jain,S. and Yuhki,N.
TITLE Molecular basis of feline beta-glucuronidase deficiency: an animal
JOURNAL Model of mucopolysaccharidosis VII
MEDLINE Genomics 58 (2), 121-128 (1999)
PUBMED 99296826
10366443
REFERENCE 2 (bases 1 to 2014)
AUTHORS Pyfe,J.C., Kurzhals,R.L. and Henchorn,P.S.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1997) Microbiology, Michigan State University,
413 Giltner Hall, East Lansing, MI 48824, USA
FEATURES
source
1..2014
location/Qualifiers
/organism="Felis catus"
/mol_type="mRNA"
/db_xref="taxon:9685"
/cell_type="fibroblast"
1..2014
/gene="GUSB"

CDS

/allele="MPS VII mutant"
2.1957
/gene="GUSB"
/allele="MPS VII mutant"
/EC_number="3.2.1.31"
/codon_start=1
/product="mutant beta-glucuronidase"
/protein_id="MAD01499.1"
/db_xref="GI:4102553"
/translation="MLRGPAAVMAALGFLMLACGLARGLMGLYRSPSPRRKELNGL
WSFRADFSERNRQGFQEWYRPLRESGPTLDMVPSPFNDVGDROLRSFVGMWWE
REATLPRMTQDLGTRVLRIGSAHYVAIWMNGVHAHEHGHLPEADISKLVOG
PLASCRITAINNTLTPHTLPGLITVOTDSKYPKGYFVNINDEFNAGLHRLPVL
LYTPPTVIDITITTSVNODTGLVDVQIPLFEGGHRDLRYRLIDDEGKVAAGTGR
GLOVNPVHMLMMPYLMEHPAYLVLSLEVRLLTAQTAAGSVSPFTLPGIRIVATHEQ
FLINCKPFPYGVNKHADIRGKFPDPLVDFPDLRLMGLAFAFRTSHPAEAVW
QLCDRYGIVIDESPGVGLVAVESYNSVLSOHLEWELVRDKNHPAVVMSSVANE
PASFLKPGAGYFKTLIAHTKALDPSRPTVTSNRYEADLGA PVVDYICVNSYSYWH
DYGHMEVILQDLATQFENMYRTYQKPIIOSGYGADTAGFHODPLMFSESYQGLER
OYHVLVDOKREYVVGELIMNFADPMTQSPQRAVGNKKGI FTTRQRPKGAAPLLRR
YMKLANETRYMNSAVKSCLENSPPTL"

ORIGIN

Query Match 11.4%; Score 192.4; DB 4; Length 2014;
Best Local Similarity 53.0%; Pred. No. 2e-40;
Matches 493; Conservative 0; Mismatches 416; Indels 21; Gaps 3;

QY 751 GACGAGTCACTCTCGGACATCGGAATCGAATCGATCAGCTGGACGAGAGAGGCTCTAT 810
DB 950 GACTTCTACCTCTCCCGTGGGATTCGACGGGTGCTGTACAGAGCACCACTTCTC 1009
QY 811 CTGAACGGAAACCTGTCTTTTGAAGGCTTTGGAAGACGAGGAATCCCGTCTG 870
DB 1010 ATCAATGGAAACCTTTATATTCACGGGGTCAACAAAGCAAGATGAGATATCCGA 1069
QY 871 GGGCAGGGCACCCTTTATTCATTGATGATTAAGACTTCAACCTTGAAGTGATCAAC 930
DB 1070 GGGAAAGGCTTGACTGGCCACTGCTGTGTAAGGACTTCAATTGCTCGTGGCTGGG 1129
QY 931 GCGAATCTCTCAGAACCTCTCATCTATCTTAAGTAAAGAGTGGCTGATCTTGGCCGAC 990
DB 1130 GCGAACGCTCTCCGACCACTGATCACTACCTTACCGGAGGAGTGATGCGAGCTTGTGAC 1189
QY 991 AGACTCGGAATCTTGTGATGAGCAAGACCCCGACCTGTGATCACAAGTATCA---C 1047
DB 1190 CGTATGGGATCGTGTGATGAGAGAGTCCCGTGTGGCANTCGTGTGTGAGAGC 1249
QY 1048 TACAATCCCGAGATCTCAGAGATAGCAGAGACAATTAAGAGATGATCGACAGACAC 1107
DB 1250 TACAGCAACGTGTCTGACACACACCTGAGAGTGTGAGAGAGCTGTGCGAGGGAC 1309
QY 1108 AAGAACATCCCACTGATGATCATGTGGAGTGTGCGAAGCAACAGAGTCCACATCCA 1167
DB 1310 AAGAAATACCCGGCTGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1369
QY 1168 GACCGGAGGCTTTCTTCAAGCCCTTTATGAGACTCCCAATGAATGATGAAACAGC 1227
DB 1370 CCTGCCGTTACTACTTCAAGACGCTGATGTGTCACACCAAGGCTTGATCCCTCCGG 1429
QY 1228 CCGGTTCATGAGAGACATGATGAGAGCAGCAGCAGAGAGAAACAGAGACGTGGCGCTG 1287
DB 1430 CCGGTGACCTTTGACCAACTCCACTATGAAG-----CAGACCTGGGGGCGC 1477
QY 1288 AAGTACTTCGACATCGTGTGTGAAACAGGTACTACGGCTGTGATCATCTATCAGGGAAAG 1347
DB 1478 CCGTATGTGACGTCATCTGTGTGAATGATTACTACTCTTGTGTATGATGCTATGTGTAC 1537
QY 1348 ATGAAGAAGAAGCTTCAAGCTCTGAAAAAAGACATAGAGAGCTCTATGCAAGGACAGA 1407
DB 1538 ATGAGAGTGAATCAGCTGACAGCTGGCAACCAAGTTGAGAACTGTATAGAGACTATCAG 1597
QY 1408 AAGCCATCTTTGTCAAGAAATTCGTTGGGAGAGCGCATAGCTGCACTACATCAATCCA 1467

DB 1598 AAACCAATTAATCCAGAGCGATGACGAGACACACCATTCGACGGGTTTCACGAGACCA 1657
QY 1468 CTTCAATGTTTCTCCGAGAGATACCAAGAGAGCTGTTGAAAAAGCATCAGGCTCCTT 1527
DB 1658 CTTGTATGTTTACGCGAGGAATACCAAGAAAGCTCTCTGAGAGAGATCATTTGTTCTG 1717
QY 1528 TTGAAA-----AAAGCTACATCATCGGAAACACAGCTGTGGCCTTTGACATTTTAAG 1581
DB 1718 GATCAAAAACGCAAAAGAAATACGTGTGCGAGAGCTCATCTGAAACTTTCGCGGATTTATG 1777
QY 1582 ACTCTCAGAAATGAGAAAGACCATTCATCAACCAAGAGGTGTTTCACAAGAGACAG 1641
DB 1778 ACTAACAGTACACCGCAGAGAGATAGGGGAATTAAGAGGATCTTACCGCCGACAG 1837
QY 1642 CAACCAAACTGTTCTCATGTACTGAGA 1671
DB 1838 CAACCAAAAGGGGACAGCTTCTTTGCGA 1867

RESULT 5

AF019759

LOCUS 2199 bp mRNA linear MAR 30-MAR-1998

DEFINITION

Canis familiaris beta-glucuronidase (GUSB) mRNA, complete cds.

ACCESSION

AF019759

VERSION

AF019759.1 GI:2425090

KEYWORDS

SOURCE

ORGANISM

Canis familiaris (dog)

Canis familiaris

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

9521879

2 (bases 1 to 2199)

Henthorn, P.S., Fyfe, J.C. and Bouvet, A.B.

Direct Submision

Submitted (15-AUG-1997)

Section of Medical Genetics, University of

Pennsylvania School of Veterinary Medicine, 3900 Delancey St.,

Philadelphia, PA 19104, USA

location/Qualifiers

1..2199

/organism="Canis familiaris"

/mol_type="mRNA"

/db_xref="taxon:9615"

1..2199

/gene="GUSB"

63..2018

/gene="GUSB"

/EC_number="3.2.1.31"

/note="lysosomal enzyme"

/codon_start=1

/product="beta-glucuronidase"

/protein_id="AAC4809.1"

/db_xref="GI:2425091"

/translation="MSRGPAGAWALGFLMTGSLALEGMLYRSPSPRRKDLGL
WSFRADFSERDQGFQEWYRPLRESGPTLDMVPSPFNDVGDROLRSFVGMWWE
REATLPRMTQDLGTRVLRIGSAHYVAIWMNGVHAHEHGHLPEADISKLVOG
PLASCRITAINNTLTPHTLPGLITVOTDSKYPKGYFVNINDEFNAGLHRLPVL
LYTPPTVIDITITTSVNODTGLVDVQIPLFEGGHRDLRYRLIDDEGKVAAGTGR
GLOVNPVHMLMMPYLMEHPAYLVLSLEVRLLTAQTAAGSVSPFTLPGIRIVATHEQ
FLINCKPFPYGVNKHADIRGKFPDPLVDFPDLRLMGLAFAFRTSHPAEAVW
QLCDRYGIVIDESPGVGLVAVESYNSVLSOHLEWELVRDKNHPAVVMSSVANE
PASFLKPGAGYFKTLIAHTKALDPSRPTVTSNRYEADLGA PVVDYICVNSYSYWH
DYGHMEVILQDLATQFENMYRTYQKPIIOSGYGADTAGFHODPLMFSESYQGLER
OYHVLVDOKREYVVGELIMNFADPMTQSPQRAVGNKKGI FTTRQRPKGAAPLLRR
YMKLANETRYMNSAVKSCLENSPPTL"

ORIGIN

AF019759

LOCUS 2199 bp mRNA linear MAR 30-MAR-1998

DEFINITION

Canis familiaris beta-glucuronidase (GUSB) mRNA, complete cds.

ACCESSION

AF019759

VERSION

AF019759.1 GI:2425090

KEYWORDS

SOURCE

ORGANISM

Canis familiaris (dog)

Canis familiaris

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

9521879

2 (bases 1 to 2199)

Henthorn, P.S., Fyfe, J.C. and Bouvet, A.B.

Direct Submision

Submitted (15-AUG-1997)

Section of Medical Genetics, University of

Pennsylvania School of Veterinary Medicine, 3900 Delancey St.,

Philadelphia, PA 19104, USA

location/Qualifiers

1..2199

/organism="Canis familiaris"

/mol_type="mRNA"

/db_xref="taxon:9615"

1..2199

/gene="GUSB"

63..2018

/gene="GUSB"

/EC_number="3.2.1.31"

/note="lysosomal enzyme"

/codon_start=1

/product="beta-glucuronidase"

/protein_id="AAC4809.1"

/db_xref="GI:2425091"

/translation="MSRGPAGAWALGFLMTGSLALEGMLYRSPSPRRKDLGL
WSFRADFSERDQGFQEWYRPLRESGPTLDMVPSPFNDVGDROLRSFVGMWWE
REATLPRMTQDLGTRVLRIGSAHYVAIWMNGVHAHEHGHLPEADISKLVOG
PLASCRITAINNTLTPHTLPGLITVOTDSKYPKGYFVNINDEFNAGLHRLPVL
LYTPPTVIDITITTSVNODTGLVDVQIPLFEGGHRDLRYRLIDDEGKVAAGTGR
GLOVNPVHMLMMPYLMEHPAYLVLSLEVRLLTAQTAAGSVSPFTLPGIRIVATHEQ
FLINCKPFPYGVNKHADIRGKFPDPLVDFPDLRLMGLAFAFRTSHPAEAVW
QLCDRYGIVIDESPGVGLVAVESYNSVLSOHLEWELVRDKNHPAVVMSSVANE
PASFLKPGAGYFKTLIAHTKALDPSRPTVTSNRYEADLGA PVVDYICVNSYSYWH
DYGHMEVILQDLATQFENMYRTYQKPIIOSGYGADTAGFHODPLMFSESYQGLER
OYHVLVDOKREYVVGELIMNFADPMTQSPQRAVGNKKGI FTTRQRPKGAAPLLRR
YMKLANETRYMNSAVKSCLENSPPTL"

		Best Local Similarity 51.6%; Pred. No. 1.8e-35; Mismatches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;	
QY	734 AGGTGAACTTGGAAAAGACGATGACACTGTGGACATCGGATCAGAACGATCAGCTGGG	793	
Db	1104 AGTTGGAATCTCTGCAGGATACCTACCGCCCTCCAGTGGGATACGATGAATGATTGGG	1163	
QY	794 ACGAGAAGGGCTCATCTGAAAGGGAAACCTGCTTTTGAAGGGCTTTGAAAGACG	853	
Db	1164 ACAAGACAGCTGTGTCTGATGATGAAAAACCTCTATCTGCGGGGATTTGACGGACG	1223	
QY	854 AGGAATTCCTCCGTTCTGGGGACGGCACTTTATTCATTGATGATTAAGACTTCAAC	913	
Db	1224 AGGATCCGATATCCGGGGAAAGGATTTGATATGGGCTTCTTGTAAGATTTTAAAC	1283	
QY	914 TTTCGAATTTGATCAACCGGAATCTTTACAGACTCTCACTTCTTACATGTAAGAT	973	
Db	1284 TGTGAAATGTGACTGGAGCCAAATGCAATGCGACCTCTCACTTCTTATTTCCGAAGAT	1343	
QY	974 GAGTGAATCTTGGCCGACAGACTCGGAATCCTTGATGATAGAGAAAGCCCGACGTTGTA	1033	
Db	1344 CAATGCAATTTGGCGATCAGCATGATGATTTATGATTAATGACGAATGCC-----TGTGT	1397	
QY	1034 TCACAGGTACCACTACATCCGAGACTCAGAAATAGCAGAAACACATAGAAGAA	1093	
Db	1398 TCAATATGATATCTTGAGCCGACGACTAATGGAGATCAATGTCTCTCGTAGAGAAC	1457	
QY	1094 TGATGCACAGACAGAAAGCCATCCGATGTGATCATGTGATGTGGCGACAGAACCG	1153	
Db	1458 TGATTCACCGGAGACAGAAACCATTCAGTGTGTTGATGTGTGGTACCAAGACCGA	1517	
QY	1154 AGTCCAAACATCCAGACGCGGAGGGTTCTTCAAAGCCCTTATGAGACTGCCAATGAA	1213	
Db	1518 GATGAAACAAGCAGGAGAGCCCTTAATACTTGAATTCCTGTAACTATGTAAAGAAA	1577	
QY	1214 TGGATCCGAACAGCGCCCGTGTGATGTGTGACATGATGACGACACAGACGAGAACAA	1273	
Db	1578 TAGCTCAGCGACGACCTCTAACCGCGCTTAAATGCCAATC-----TTTCCA	1625	
QY	1274 GAGACGTGGCGCTGAATGACTTGCACATCGTCTGTGAAACAGTACTACGCTGGTACA	1333	
Db	1626 GTTGCATTTTGGGCGAATTTCTGACATCGTGGGTTTCAATCGCTATTACTCTTGGTATC	1685	
QY	1334 TCTATCAGGGAAGATGAAAGAACTTCAGCTCTGAAAAAAGCATAGAAGACTCT	1393	
Db	1686 AGAATCGGGCGGTACATGATGATCTTAACTGTTGACATAGAGGCCACAGAGTTGAC	1745	
QY	1394 ATGCAGAGCAAGAAAGCCCATGTTTGTCAACAAATTCGGTGGCGAGGATAGCTGGCA	1453	
Db	1746 GGGATAGGTTCCGAAGACCTGTGATCCAAATTTGATAGCGCGCGACACTATAGAGGGCA	1805	
QY	1454 TCCACTACGATCCACTCAATGTTCTCCGAAGAGTACCAGAGAGCTCGT-----	1505	
Db	1806 TGCACTCACTTCCCGCTTATTTTGGTCGGAGGAATACAGAGTTGAGCTCTTCTCGCGC	1865	
QY	1506 -TGAAAAGAGATCAGGCTCCTTTGAAAAAAGACTACATCATCGGAAACACAGTGTGGG	1564	
Db	1866 ATTTCAGAGGCTTTCGACAGATTACGGGAGAGAGATGTTTATTTGAGAGTTTGTTTGGA	1925	
QY	1565 CCTTTCAGATTTTAAAGACTCCTCAGAAATGTGAGAAAGCCCATTTCCACACCAAGGGTG	1624	
Db	1926 ACTTCGGCGAATTTTCGAGCGCGGACATATTAACCGCGTGGCGGACAAAAAAAGGAG	1985	
QY	1625 TTTTCACAGAGACAGCAACCCAAATCTGTTGCTCATGTACTGAGAAAGCTGT	1678	
Db	1986 TCTTTTACAGGAACCGACAAACCCAAAGAGTGGCTCACTTCTTGGCGCGCGGT	2039	
RESULT 8			
LOCUS BC014142			
DEFINITION BC014142 2182 bp mRNA linear PRI 29-JUN-2004			
IMAGE:4662011, complete cde.			
Homo sapiens glucuronidase, beta, mRNA (CDNA clone MGC:20659			

ACCESSION	BC014142
VERSION	BC014142.2
KEYWORDS	MG
SOURCE	MG
ORGANISM	Homo sapiens (human)
REFERENCE	1 (bases 1 to 2182)
AUTHORS	Strauberg, R.L., Fellingold, E.A., Grouse, L.H., Derge, J.G., Klusner, R.D., Collins, F.S., Wagner, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buecaw, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Ditchenko, L., Mardina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stepleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Schmitt, T.E., Brownstein, M.J., Ueda, T.B., Toshiyuki, S., Carinini, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulik, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Paley, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywinski, M.I., Skalek, A., Smalls, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
REFERENCE	2 (bases 1 to 2182)
AUTHORS	Strauberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (10-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT	On Dec 19, 2003 this sequence version replaced gi:1559559.
	Contact: MGC help desk
	Email: cgapbs-remail.nih.gov
	Tissue Procurement: ATCC
	cDNA Library Preparation: Rubin laboratory
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
	Gaithersburg, Maryland;
	Web site: http://www.nisc.nih.gov
	Contact: nisc.mgc@nih.gov
	Akhter, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granito, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legsby, R., Madhusu, Q.L., Masetto, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McEwell, J., Pearson, R., Stancijop, S., Thomas, P.J., Touchman, J.W., Tsoung, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
SOURCE	Series: REAL Plate: 30 Row: o Column: 15
	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504222.
	Location/Qualifiers
	1..2182
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="MGC:20659 IMAGE:4662011"
	/cfeature_type="Colon, adenocarcinoma"
	/clone_lib="NIH MGC_15"
	/lab_host="DH10B-R"
	/note="Vector: pOTB7"

gene	1..2182
	/gene="GUSB"
	/note="synonym: MPS7"
	/db_xref="LocusID:2990"
	/db_xref="MIM:253220"
CDS	2..1957
	/gene="GUSB"
	/codon_start=1
	/product="glucuronidase, beta"
	/protein_id="AAH14142.1"
	/db_xref="GI:1559560"
	/db_xref="LocusID:2990"
	/db_xref="MIM:253220"
	/translation="MARSAVMAALAPRLMGCAILOGGMLTYQESPRECKELDGLWSFADPSDNRRRGEQWRRPIMESGPTVDMPEVPSFNDISQWRLLHFVGMWYREVLIPERMTODLRTVRLVIRGSASVYAIWVNGVDLLEHGGLPEFADI.SNLVQVGPLPRLRTTIAINNTLPTTLPPTGIQVLTDSKPKYFYQNTYDFFNAGLQRSVLLYPTPTVYIDITVTTSSVEODSGLVNIVSKSNLPEKLEVLRLDAENKVVANGTQGLKVPVSLMWVPMHWRPAVLYSLVQLTQTSIGPVSDFTLTPVGTIVATKSOPLNGRPFYFHYGNKEDADIIRGKPFMPLIVDPRILRLGANAFTSHYPAEVMQMDRGTGVIDECPGVGLLPFPNNVSLHHMQVEVVRKDKHPAVVMSVANEPASHLBSAGYLLKMLVIAHTSLDPSRPFVSNSNAADGAPVVDVCLNSYVSWYHDYGLLELIQIQLATQFENYKTKQKPIIQEYGAETIAGHODPPLMFTEYQKSLLEYHGLDQKRKYVVGELIWNFADPMTQSPTRVLGNKGIIFTRQRPKSAALFLRERYWKLINETRYPHVAVSAQCLNENLFT"
ORIGIN	
Query Match	10.3%; Score 174; DB 9; Length 2182;
Best Local Similarity	52.2%; Pred. No. 1.8e-35;
Matches	493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
QY	737 TGAACCTTGAAGAAAGAGAGTACCTCTGACATCGGAATCAGAACATCAGCTGGAGC 796
DB	939 TGGGCGCTGTCTGACCTTCAACACACCTCTGGGATCGCAGCTGCTGACCA 998
QY	797 AGAAGAGCTCTATCTGAACGGAAACCTGCTTTTGAAGGCTTGAAGCAGAGC 856
DB	999 AGAGCACTTCTCTATCAATGAGAAACCTTCTATTTTCAAGGTCAACAGCATGAGG 1058
QY	857 AATTCCCGCTTGGGGCAGGGCACCTTTTATTCATGATGATTAAGACCTTCAACCTTC 916
DB	1059 ATGCGCATCCGAGGAGGAGGCTTGCAGCTGGCGCTCTGTGAAGACCTTCAACCTGC 1118
QY	917 TGAAGTGAATCAAGCCGAATTTTCAAGACCTTCAATATCTTCAAGTGAAGAGTGGC 976
DB	1119 TTCCGTGCTGTTGTGCCAACCTTCCGTACAGCACTACCCCTATCAGAGGAGTGA 1178
QY	977 TGGATCTTGGCAGACAGATCGGAATCCCTTGATGATGAGAAAGCCCGACCTGGTATCA 1036
DB	1179 TGCAGATGTGACCGCTATGAGATTTGTGATGATGAGTGTCCGGCGTGGCTTGG 1238
QY	1037 CAAGGTACCA--CTACATCCCGAGACTCAGAGATGAGAGAAAGCAATTAAGAGAA 1093
DB	1239 CGCTGCGCAGTTCTTGAACAAGTTTCTTGACATCACCAGATGAGGTGATGAGAAAG 1298
QY	1094 TGATGACAGACACAGAACATCCCACTGTGATCATGTGAGTGTGGCGAAGCAAG 1153
DB	1299 TGTGTCGTGAGGACAGAAACACCCCGGTGTGATGTGTGTGTCGTCACCAAGGCTTG 1358
QY	1154 AGTCCAACTTCAGAGCGGAGGCTTCTTCAAGACCTTTATGAGAGCTGCATATGAA 1213
DB	1359 CGTCCACCTTGAATCTGCTGCTACTACTTGAAGANTGTATGCTTCAACCAATCTCT 1418
QY	1214 TGATGGAACAGCGCCCTGTGTCACTGTGACATGATGAGCAGCAGAGAGAGAA 1273
DB	1419 TGAACCCCTCCGGCGTGTGACCTTTGTGACCACTGAATGAGCAAG----- 1466
QY	1274 GAGAGTGGCCCTGAAGTCTTGCACATGCTGTGTGAACAAGTACTAGCGCTGTCA 1333
DB	1467 CAGACAGAGGGGCTCCGTATGTGATGATGATCTGTGAAACAGTACTACTCTTGTATC 1526
QY	1334 TCTATCAGGAAGATGAGAAAGAGCTTCAAGCTCTGGAAGAAAGACATAGAGAGCTCT 1393

Db 1527 ACGACTACGGGACACCTGGAGTTGATTCAGCTGACGCTGGCCACCCAGTTTGAAGACTGCT 1586
 Qy 1394 ATGCAAGGACAGAAAGCCCTCTTTGTGACAGAAATTCGGTGGAGCCGATAGCTGGCA 1453
 Db 1587 ATAGAAGTATCAGAAAGCCCATTTATTCAGACCGAGTATGAGCAGAAACGATTCAGAGGT 1646
 Qy 1454 TCACACTACGACCCACCTCAATGTTCTCCGAAAGTACCAAGCAGAGCTGGTGAAG- 1512
 Db 1647 TTCACAGAGATCCACCTTGTATGTTCACTGAAAGATACAAAGATCTGCTAGAGCAT 1706
 Qy 1513 ACGATCAGGCTCTTTTGAAGAA-AGACTATCATCTGGAACACACGCTGTGGCCT 1567
 Db 1707 ACCATCTGGTCTGGATCAAAAGCAGAAATATCGTGGTTGAGAGCTCATTTGAAATT 1766
 Qy 1568 TTGCAATTTTAAAGCTCTCAGATGTGAAAGACCATTTCTACCAAGAGGTGTTT 1627
 Db 1767 TTGGCGATTTATCATCTGAACAGTACCGCAGAAAGTGGGAAATMAAAGGGGATCT 1826
 Qy 1628 TCACAGAGACAGACCAACCACTGTTGCTCATGTACTGAGA 1671
 Db 1827 TCATCGGACAGACACCAAAAGTGCAGCGCTTCTTTGCGA 1870

RESULT 9
 BT010073 2321 bp mRNA linear INV 12-AUG-2003
 LOCUS Drosophila melanogaster LD10588 full insert cDNA.
 ACCESSION BT010073
 VERSION BT010073.1 GI:33589549
 KEYWORDS FL1 CDNA.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 2321)
 Stadelman, M., Brokstein, P., Hong, L., Aghayani, A., Carlson, J.,
 Champagne, R., Chavez, C., Doresek, V., Dresnek, D., Farfan, D., Fiske, E.,
 Giarra, M., Gonzalez, M., Guarin, H., Krommiller, B., Li, P., Liao, G.,
 Miranda, A., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
 Patel, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M.
 and Celisner, S.
 Direct Submission
 Submitted (12-AUG-2003) Berkeley Drosophila Genome Project,
 Lawrence Berkeley National Laboratory, One Cyclotron Road,
 Berkeley, CA 94720, USA
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory
 Berkeley, CA 94720

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 This clone was sequenced as part of a high-throughput process to
 sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
 Science 2000). The sequence has been subjected to integrity checks
 for sequence accuracy, presence of a polyA tail and contiguity
 within 100 kb in the genome. Thus we believe the sequence to
 reflect accurately this particular cDNA clone. However, there are
 artifacts associated with the generation of cDNA clones that may
 have not been detected in our initial analyses such as internal
 priming, priming from contaminating genomic DNA, retained introns
 due to reverse transcription errors of unspliced precursor RNAs, and
 reverse transcriptase errors that result in single base changes.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our Web site
 (http://fruitfly.berkeley.edu) or send email to
 cda@fruitfly.berkeley.edu.
 Location/Qualifiers
 1..2321
 /organism="Drosophila melanogaster"
 /mol_type="mRNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 1..2321
 /gene="CG2135"

FEATURES

source

gene

CDS

/db_xref="FLYBASE:FBgn0039874"
 64..2124
 /gene="CG2135"
 /note="Longest ORF"
 /codon_start=1
 /product="LD10588p"
 /protein_id="AA02542.1"
 /db_xref="GI:33589550"
 /db_xref="FLYBASE:FBgn0039874"
 /translation="MRSHMPVYCLLGLQVLAVTENDFLVNRKELTDSPLFEY
 DNRKPTPKGLLYPRDSSETRSLDGMQLRSDYDPLQIRKEMDLARKGR
 EISMVPASVNDITTDNLRDHYGTWYERTFVPSKMTORLRFSSVSAVW
 INRNATSHSIGLIPRESISGLISFGGNRLITVWCDNLSNRITPOGSAVYKATDQ
 TYPIDSTPDPFRVNAIHRSHVLYTTPPLHISLSEVYTLQTEGLRIDYRVLDLNSK
 BGLQIPIQLRQVLRDGHVAAQIQINKVATHTDLVPAATFWPPLMHSDPEYLYNL
 QFELFVANSKEKESLDQTYRLPVGIRLSMDNDSLLNGKPLYLFGFGRHSDIRG
 KGLDNLALRDPLILKMTGANAYRTSHYSEBSMQPADOHIMIDECPAVVIDE
 POLLENHMSLEQLIHRDNRHPSVVAASVANSRSMKQALKYFELVAVYRIRAHG
 PLTAANANSSCHLAQPLDIVFNRYNSWYONSRTDMLINLTTEAQSMDRFPKRP
 VIOFEYGDPMEGMSHLPAPIWSEYQVLFPSHFPAPDELRGRGFIQEPVNPADF
 RTRQITTRVGANKGVFTRNRQPEVAHILRRKYPALAKELDMFGLPDLVTIISKHI
 HNEL"

ORIGIN

Query Match 10.3%; Score 174; DB 3; Length 2321;
 Best Local Similarity 51.6%; Pred. No. 1.8e-35;
 Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;

Qy 734 AGGTGAACTTGAAGAAAGACGAGTACATCTGACATCGGAATCAGAACATCAGCTGGG 793
 Db 1100 AGTGGAACTCTGACAGATACCTACCGCTCCAGTGGGCAATAGCTTAAGTTGGG 1159
 Qy 794 ACGAAGAGAGGCTCTATCTGGAAGGGAACCTGCTTTTGAAGGCTTTGGAAGACAG 853
 Db 1160 ACGAAGAGAGGCTCTGCTGGAAGGGAACCCCTCATCTGCGGGATTTGACGCGACG 1219
 Qy 854 AGGAATTTCCCGTCTGGGGACAGGACCTTTTATCATTTGATGATAAAGACTTCAACC 913
 Db 1220 AGGATCTCGATATCCCGGGAAGGATTTGATATCGCTTCTGCTAGAGATTTTAAC 1279
 Qy 914 TTCTGAAGTGATCAACGCGGAATTTTTCAGAGCTCTGATACCTTTCAGTGAAGAGT 973
 Db 1280 TGCTGAAGTGAGATGAGCAATGATATGACCTCTCATATCTTATTCGGAAGAGT 1339
 Qy 974 GGTGATCTTGGCCGACAGCTCGGAATCTTGTGTATACGAAGCCCGGACGTTGTA 1033
 Db 1340 CAATGCAATTTGCGGATCAGCATGTATTATGATTAACGATGCC-----TGCTG 1393
 Qy 1034 TCACAAGTATACCACTCAATCCGAGACTCAGAAGATAGCAGAAACAATTAAGAAGAA 1093
 Db 1394 TCAATATGATATCTTCGAGCCGACGCTACTGGAAGATCAGATGCTCTGGCTGAGCAAC 1453
 Qy 1094 TGATGACAGACACAAAGAACATCCAGTGTATCATGTGAGTGTGGCAACGAACAG 1153
 Db 1454 TGATCCACCGGACAGAAACATCCAAAGTGTGTCATGTCGTGAGGCAACGAGCCCA 1513
 Qy 1154 AGTCCAAACATCAGACGCGGAGGAGGTTTCTTCAAGCCCTTATAGACTGCCAATGAAA 1213
 Db 1514 GATCGAACAAGACGAGGAGCCCTTAATTAATCTTGAATCTTGTAATCTATGTAAGAAA 1573
 Qy 1214 TGGATGAAACAGCCCGCTGTGTATGATGAGATGAGACGACGACGAGAGAACA 1273
 Db 1574 TACGTACGAGGAGACCTTAACCGGGCGTATTAATGCCAACTC-----TTCCA 1621
 Qy 1274 GAGACGTGCGCTGAAGTACTTGACATGCTGTGTGAACAGTACTACGGCTGTACA 1333
 Db 1622 GTTGCAATTTGGCGCGCTTTCGACATGTGGGTTCATGCTATTAATCTGTTGATAC 1681
 Qy 1334 TCATATGAGGAGATAGAGAAGAGACTTCAAGCTCTGGAAGAAAGCATAGAAAGCTCT 1393
 Db 1682 AGAATCTGGGCGGTATGATGATATCTTAACCTGTTGACGATAGAGGCCAGAGTTGCG 1741
 Qy 1394 ATGCAAGGACAGAAAGCCCATCTTTGTACAGAAATTCGATCGGACGAGATAGCTGGA 1453

```

Db      1742 GGGAGAGGTGGGAAAGCTGTCTATCAATTGATAGCGCGGCACTATGAGGGCA 1801
Qy      1454 TCCACTACGATCCACTCAATGTTCTCCGAAGAGTACCAAGACGCTCGT----- 1505
Db      1802 TGCATCTCACTTCCGCCCTTTATTTGTCGGAGAAATACAGAGTTGAGCTCTCTCGCC 1861
Qy      1506 -TGAAAGACGATCAGGCTCCCTTTGMAAAAGACTACATCATCGGAACACAGTGTGG 1564
Db      1862 ATTCAAGGCTTTCAACGAGTTACGCGGAGAGAGTGTATTGTGAAGATTGTGGA 1921
Qy      1565 CTTTGACAGATTTTAAGACTCCTCAGAAATGTGAGAAGACCATTTCTCAACCAAGGTTG 1624
Db      1922 ACTTGCGGATTTTGGACCGGCGAGACTATTAACCGGTGGCGGCAAAAAGGAG 1981
Qy      1625 TTTTCAACAAGACACAAACCAACCTCGTTGCTCATGTACTGAAAGACTGT 1678
Db      1982 TCTTTACAAGAACCGAACCAACCAAGAGTGTCTCATCTTTAGGCGGCGGT 2035

```

RESULT 10

```

AR344264      1956 bp      DNA      linear      PAT 17-AUG-2003
LOCUS      AR344264
DEFINITION      Sequence 1 from patent US 6582692.
ACCESSION      AR344264
VERSION      AR344264.1 GI:33740191
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 1956)
AUTHORS      Podsakoff, G. and Watson, G.
TITLE      Recombinant adeno-associated virus virions for the treatment of
JOURNAL      Patent: US 6582692-A 1 24-JUN-2003;
FEATURES
source      Location/Qualifiers
1..1956
/organism="Unknown"
/mol_type="genomic DNA"

```

```

ORIGIN
Query Match      10.2%; Score 172.4; DB 6; Length 1956;
Best Local Similarity 52.1%; Pred. No. 4,9e-35;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

```

Qy      737 TGGAACTTGAAGAGCAGATCACTTGGACATCGAATCAAGATCAAGCTTGGAG 796
Db      938 TGGGGCTGTGCTGCTGCTACACATCCCTGTGGGATCCGACTGTGCTGCACCA 997
Qy      797 AGAAGAGCTCTATCTGAACCGGAACTGTCTTTTGAAGGCTTTGMAAGCAGAG 856
Db      998 AGAGCAATTTCTCTCATATGGAACCTTTCTATTTCCACGCTTCAACAAAGCATYAGG 1057
Qy      857 AATTCCTCGTCTGGGGGAGGCGACCTTTTATCATGATGATTAAGAAGCTTCACTTC 916
Db      1058 ATGCGGAATCCGAGGAGAGGCTTGCATGCGCTCTGTGTGAAGACTTCAACTGC 1117
Qy      917 TGAAGTGATCAACGCAATTTCTTCAAGACCTTCACTATCTTCAAGTGAAGTGC 976
Db      1118 TTGCTGTGCTGTGTCGAACGCTTTCCGTACAGCACTAACCTTATGCAAGAGTGA 1177
Qy      977 TGAATCTTGGCCGACAGACTCGGAATCTTGTGATGAGAGAGCCCCGACGTTGTATCA 1036
Db      1178 TGCAGATGTGTGACCGCTATGGGATTTGTGTATGATGATGATGTCGCGCGGTGG 1237
Qy      1037 CAAGGTACCA--CTACATCCCGAGACTCAGAAATGAGAGAAACAATTAAGAA 1093
Db      1238 CGCTGCCGAGTTCTTCAACAGCTTTCTTCTGATCAACACATGCAAGTATGAGAAAG 1297
Qy      1094 TGAATGACAGACAAAGAACATCCAGTGTATCATGTGAGTGTGCGAAGCAACGAG 1153
Db      1298 TGTGTGTAAGGAGCAAGAACCAACCCGCGGTGTGTATGTGTGTGTGTGCGAAGAGCTGT 1357

```

```

Qy      1154 AGTCAACCATTCAGAGCGAGGGTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA 1213
Db      1358 GTTCCACCTTGAATCTGTGCTACTACTTGAAGATGTATGTCTACACCAATCTCT 1417
Qy      1214 TGAATGAAACAGCGCCCTTGTCTATGTGAGACATGATGAGACGACAGAGAAACAA 1273
Db      1418 TGAACCTCTCCCGGCTGTGACTTTGTGACCACTCACTATGACAG----- 1465
Qy      1274 GAGAGTGGCGCTGAAGTACTTGCATGCTGTGTGTAACAAGTACTAGCGCTGTACA 1333
Db      1466 CAGACAAAGGGGCTCCGTATGTGATGTGATCTTTTGAACAGCTACTACTCTTGTATC 1525
Qy      1334 TCTATCAGGAAGATGAGAAGAGCTTCAAGCTCTGGAAGAAAGACATAGAGCTCT 1393
Db      1526 AGACTACGGGCACTGAGATTGATTCAGCTGCAAGCTGCGCACCACTTGAAGCTGT 1585
Qy      1394 ATGCAAGGACAGAAAGCCATCTTTTGCACAGAAATGCTGCGGACCGATAGCTGCA 1453
Db      1586 ATTAAGAGTATCAGAAAGCCATTTATTCAGAGGAGTATGAGACAGAAACGATTCAGAGGT 1645
Qy      1454 TCCACTACGATCCACTCAATGTTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512
Db      1646 TTCAACAGAGATCACTCTGATGTTCAGTGAAGAGTACCAAGAAAGTCTGTAGAGCAGT 1705
Qy      1513 AGCATCAGGCTCTTTTGAAGAA-----AGACTACATGATGGAACACAGGTGGCGCT 1567
Db      1706 ACCATCTGGGCTGTGATTAACGAGAAATATGTGTGAGAGAGTCAATTGGAATT 1765
Qy      1568 TTGCAATTTTAAGACTCCTCAGATGTGAGAAGACCATTTCTCAACCAAGGCTGTTT 1627
Db      1766 TTGCGGATTTTATGATCTGAACAGTACCGGAGAGGTCTGGGGAATTAAGAGGAGACT 1825
Qy      1628 TCACAGAGACAGACAAACCAACCTGTTGCTCATGTACTGAGA 1671
Db      1826 TCATCGGACAGACAAACCAAGAGTGCAGCGTTCCTTTTCGA 1869

```

RESULT 11

```

AX147652      1956 bp      DNA      linear      PAT 08-JUN-2001
LOCUS      AX147652
DEFINITION      Sequence 1 from Patent W00136603.
ACCESSION      AX147652
VERSION      AX147652.1 GI:14346708
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1
AUTHORS      Podsakoff, G., Watson, G., Couto, L.B. and Yang, B.
TITLE      Recombinant adeno-associated virus virions for the treatment of
JOURNAL      Patent: WO 0136603-A 1 25-MAY-2001;
FEATURES
source      Location/Qualifiers
1..1956
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
1..1956
/note="unamed protein product"
/codon_start=1
/protein_id="CAC41170.1"
/db_xref="GI:14346709"
/translation="MARGSAVAMALGLPILMGALGLOGGLVPOBSPRECKELDGL
MSPADPSDNRBGRFEROWRRLPLWESGAPYCDWBPSSPNDISQDMLRHVGWVYE
REVLPKRWTDLTRVVLKRGSHSAIYVWNCVDTLHEHGDTLPPEADI SNLVYOG
PLPSRLITIAINNTLPTLPPTLPGCTIYLDTSTKYPKGYPVQNTYFDFPNVAGLQSV
LLVLTPTVYIDDIIVTTSVDSGLVNYOISVKSXNLFKLEVRLLDENKRVANGTGT
GQGLKVPVGSILMPYLMHBRPAYLYSEVOLTQTSIGPVSDFYTLPEVLRIVAVTGS
OFLINGRPYFHVGNKIEDIDRGKFDWPLVLDPNLRLWLGANAPRTSHYPAEEV
MOMCDRGIIVIDCPGVGLALPFPFNVLVLIHNMQMEVVRDRDKXHPAVVMVSVAN

```

CDS

```

1..1956
/note="unamed protein product"
/codon_start=1
/protein_id="CAC41170.1"
/db_xref="GI:14346709"
/translation="MARGSAVAMALGLPILMGALGLOGGLVPOBSPRECKELDGL
MSPADPSDNRBGRFEROWRRLPLWESGAPYCDWBPSSPNDISQDMLRHVGWVYE
REVLPKRWTDLTRVVLKRGSHSAIYVWNCVDTLHEHGDTLPPEADI SNLVYOG
PLPSRLITIAINNTLPTLPPTLPGCTIYLDTSTKYPKGYPVQNTYFDFPNVAGLQSV
LLVLTPTVYIDDIIVTTSVDSGLVNYOISVKSXNLFKLEVRLLDENKRVANGTGT
GQGLKVPVGSILMPYLMHBRPAYLYSEVOLTQTSIGPVSDFYTLPEVLRIVAVTGS
OFLINGRPYFHVGNKIEDIDRGKFDWPLVLDPNLRLWLGANAPRTSHYPAEEV
MOMCDRGIIVIDCPGVGLALPFPFNVLVLIHNMQMEVVRDRDKXHPAVVMVSVAN

```

EPASHLESAGYLLKMWIAHTXSLDPSRPTFVNSNVAADKAPYVDVICINLSYMY
HDYHGLIELIOLATOFENMYKYOKPIIOSEVCAETIAGFHOPPLMFREYKSLI
EYHGLDOKRKRYVSELINWPRADPMTIOEPTVLGNKKIIFTROROPSAAPLRE
RWKIANETRIPHSAKSCLENSPFT"

ORIGIN

Query Match 10.2%; Score 172.4; DB 6; Length 1956;
Best Local Similarity 52.1%; Pred. No. 4.9e-35;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

Db 737 TGGAACTTGAAMAAACAGATACCTCTGAGCATCGAATGAGAACATGATGAGTGGAGCG 796
Db 938 TGGGGCTGTGTCTCTTACACATCCCTGTGGGATCCGACCTGTGCTGTACCA 997
Qy 797 AGAAGAGCTCATCTGAACGGGAACTGTCTTTTGAAGGCTTTGGAAGCAGAGG 856
Db 998 AGAGCAAGTTCTCTATCATAGGAAACCTTTCTATTTCCAGGGTGTCAACAGATAGAG 1057
Qy 857 AATTCCCGCTTCTGGGGCAGGGCACCTTTTATCCATGATGATAAAGACTTCAACCTTC 916
Db 1058 ATCCGACATCCGAGGGAAGGGCTTCGACTGGCCGCTCTGTGAAGACTTCAACCTTC 1117
Qy 917 TGAAGTGAATCAACCGCAATCTTTCAAGACCTCTCATATCTTTACAGTGAAGAGTGC 976
Db 1118 TTCCGTGCTTGTGTCACACGCTTCCGTACACGCACTACCCCTATGCAAGAGATGA 1177
Qy 977 TGGATCTTGCCGACAGACTCGGAATCCTGTGATGAGAGAGCCCGACCTTGTGTCA 1036
Db 1178 TGCAGATGTGTACCGCTATGAGATGTGTGATCATGATGATGTGTCGGGCTGTGGCTGG 1237
Qy 1037 CAAGGTACCA--CTACATCCGAGACTCAGAAATAGCAGAAAGACAACATTAAGAGAA 1093
Db 1238 CGTGCAGCACTTCTTGAACACGTTTCTGCAATCAACATGCAAGGTATGGAAGAG 1297
Qy 1094 TGAATGACAGACACAGAACCATCCCACTGTGATCATGTGAGTGTGGCGAAGCAACGAG 1153
Db 1298 TGTGTGCGTGAAGAGAACACACCCCGGCTGTGATGTGTGTGTGTGCGCAACGAGCTG 1357
Qy 1154 AGTCCAAACATTCAGACGCGGAGGGTTTCTTCAAGGCCCTTATGAGACTGCGCAATGAA 1213
Db 1358 CGTCCACACTTGAATCTGTCTGTGCTAATCTTGAAGATGATGATGCTCACACCAATCTT 1417
Qy 1214 TGAATGACACACGCGCGTGTGTCAATGTGAGACATGATGAGCAGCAGAGAGAGAA 1273
Db 1418 TGAACCCCTCCCGGCTGTGACCTTTGTGAGCACTAATCTATGAG----- 1465
Qy 1274 GAGAGCTGCGCTGAAGTACTTCAATCTGTGTGTGAGACAGTACTAGCGCTGTGACA 1333
Db 1466 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTGTAACAGCTACTACTTGTGTATC 1525
Qy 1334 TCTATCAGGGAAGATAGAAAGAACTTCAAGCTTGAAGAAAGACATTAAGAGAGCTCT 1393
Db 1526 ACGACTACGGGCACTGGAAGTGAATTCAGCTGCGAGCTGGCCACCACTTTGAAGACTGCT 1585
Qy 1394 ATGCAAGGACAGAAAGCCATCTTTGTCAAGAAATTCGCTGCGAGCGATAGCTTGCA 1453
Db 1586 ATAAAGAGTATCAGAAAGCCATTAATCAGAGCGAGTATGAGAGCAAGAAACGATTGCA 1445
Qy 1454 TCCACTACGATCCCACTCAATATTTCTCCAGAGGTAACAGCAGAGCTGTTGAAGAG- 1512
Db 1646 TTCCACAGATCCACCTCTGTATGTTCATGAAGATACAGAAAGTCTCTGAGAGCAGT 1705
Qy 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGGAACACAGTGTGGGCTT 1567
Db 1706 ACGATCTGGGTCTGGAATCAAAAACAGAAATATGTGTGGAAGAGCTATTTGAATTT 1765
Qy 1568 TTGCAAGATTTAAGATCTCTCAGAAATGAGAAAGCCATTTCAACACAGAGGTGTTT 1627
Db 1766 TTGCGGATTTTCATATGAAACAGTCAACGACGAGAGTGTGGGAAATTAAGAGGAGTCT 1825
Qy 1628 TCACAAAGACAGACCAACCAAACTCGTTTCTATGTACTGACGAA 1671
Db 1826 TCACCTCGGACAGACCAACCAAAAGTGCAGCGTTCCTTTTGGCA 1869

RESULT 12

HUMGLCB

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

HUMGLCB 2191 bp mRNA linear PRI 08-NOV-1994
Human beta-glucuronidase mRNA, complete cds.
M15182.1 GI:183232
alternative splicing; beta-D-glucuronoside glucuronosylhydrolase;
beta-glucuronidase; glucuronidase; glucuronohydrolase; hydrolase.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2191)
Oshima,A., Kyle,J.W., Miller,R.D., Hoffmann,J.W., Powell,P.P.,
Grubb,J.H., Sly,W.S., Tropak,M., Guise,K.S. and Gravel,R.A.
Cloning, sequencing, and expression of cDNA for human
beta-glucuronidase
Proc. Natl. Acad. Sci. U.S.A. 84 (3), 685-689 (1987)
87118233
3468507
Original source text: Human fibroblasts cDNA clone pHUGF and
placenta cDNA clones pHUGP1 and pHUGP15.
Draft entry and copy of computer-readable sequence of [1] kindly
provided by R.D. Miller, 08-MAY-1987.
Four potential N-linked glycosylation sites are located at bases
543-551, 840-848, 1284-1292 and 1917-1925. Potential poly-A
signals were found 25 and 48 bp upstream of the poly-A site at
nucleotides 2139-2144 and 2162-2167 respectively.
Two types of mRNA exit due to alternative splicing. The short form,
lacking bases 939-1091 below, encodes a protein lacking
beta-glucuronidase activity.
Location/Qualifiers
1..2191
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="7q22"
1..2191
/gene="GUSB"
/gene="GUSB"
/gene="GUSB"
/product="glucuronidase-b mRNA"
27..1982
/gene="GUSB"
/note="beta-glucuronidase precursor (BC 3.2.1.31)"
/codon_start=1
/protein_id="AA52561.1"
/db_xref="GI:183233"
/db_xref="GDB:600-120-025"
/translation="MARSAVAMALAGPLNGCALGLOGMILYPOESPSRBECKEIDGL
WSFRADFSNRRRGEQWYRRPPIESGPTVDMPPSSFNDSQMRLEHFGWWE
REVILPERWTDLRTRVLRIGSAHSVAIWNVGDTLEHSEGYLPFEADISNLQV
PLPSRLRTIALINNTLPTLPPTGPIQVLTGSKYKPYFVONTYDFENYGLORSV
LVTPTPTIIDIITVTSVEDSGLVNTOISVGSNLPRLEVRILDAENKVVANNGT
OGSLKVPQVYLMWPIAMERPAVLYLSLEVLTAQSLSPVSPFTLPLPGICRVATKS
OPLNGKPPYRFGVAKHEDADIRGKGFMPFLVILKDFNLRLGANAFTSHYPAEV
NOMCDRGIYVIDECVGLALPOEFNNVSLHHNQWVEVVRDRKDHAPVAVMSVAN
EPASHLESAGYLLKMWIAHTXSLDPSRPTFVNSNVAADKAPYVDVICINLSYMY
HDYHGLIELIOLATOFENMYKYOKPIIOSEYCAETIAGFHOPPLMFREYKSLI
EYHGLDOKRKRYVSELINWPRADPMTIOEPTVLGNKKIIFTROROPSAAPLRE
RWKIANETRIPHSAKSCLENSPFT"

FEATURES

source

gene

mRNA

CDS

sig_peptide

mat_peptide

ORIGIN

Query Match 10.2%; Score 172.4; DB 9; Length 2191;
Best Local Similarity 52.1%; Pred. No. 4.9e-35;

Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

Qy 737 TGGAACTTGGAAAAAGCACTCTGAGCATGGGAATCGAAACGATCAGCTGGAGC 796
Db 964 TGGGGCTGTGTCTGACTTCTACACATCCCTGTGGGGATCCGCACTGGCTGTACCA 1023
Qy 797 AGAAGAGCTCTATCTGAAACGGGAAACCTGCTTTTGAAGGGCTTTGGAAACGAG 856
Db 1024 AGAGCAAGTCTCTCATCAATGGGAAACCTTTCTATTTCCACGGGTCAACAAAGCATGAGG 1083
Qy 857 AATTCCTCCGTTCTGGGGGAGGACCTTTTATCATGATGATTAAGACTTCAACTTC 916
Db 1084 ATGCGACATCCGAGGAAAGGCTTGCAGTGGCCCGCTGTGTGAAGACTTCAACTTC 1143
Qy 917 TGAATGATCAACGGGAATTTTTCAGACCTTCACTATCTTCACTGATGAAGTGGC 976
Db 1144 TTCCTGCTTGTGTCCCAACGCTTTCCTGACAGCACATCCCTATGACAGGAAATGA 1203
Qy 977 TGGATCTTGGCCGAGACTCGGAATCTTGTGATGAGAGAGCCCGGACGTTGATCA 1036
Db 1204 TGCAGATGTGTGACCGCTATGGGATTTGTGTATGATGATGATCCGGCGTGGCCTGG 1263
Qy 1037 CAAGGTACCA---CTACATCCCGAGACTCAGAGATAGCAGAGACAACATAGAGA 1093
Db 1264 CGCTCCGCGAGTTCTTCAACAAAGTTTCTGTGCATCACACATGCAAGGTATGAGAAG 1323
Qy 1094 TGATGACAGACACAAAGCAATCCCAAGTGTATCATGTGAGTGTGGCAACGAC 1153
Db 1324 TGTGTCGATGAGGACAAAGCAACCCCGGCTGTGTATGTGTGTGGCAACGAGCTG 1383
Qy 1154 AGTCAACCATCCAGACGGGGGTTTCTTCAAGCCCTTATAGAGATCGCCATGAAA 1213
Db 1384 CGTCCCACTAGAACTGTGTGCTACTACTTAAAGTGTATCCCTCACACCAATCCT 1443
Qy 1214 TGGATGCAACAGCCCGCTTGTCTATGATGAGCATATGAGCAGCAGAGAGACA 1273
Db 1444 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAACATGACAG----- 1491
Qy 1274 GAGAGCTGGCGCTGAAGTACTTCAACATGCTGTGTGACAGGTACTACGGCTGTACA 1333
Db 1492 CAGACAGGGGCTCCGATGTGATGTGATCTGTTGAAACAGTACTACTTGTGTATC 1551
Qy 1334 TCTATCAGGGAAGATAGAGAAAGACTTCAAGCTCTGSAAGAAAGACATAGAGAGCT 1393
Db 1552 AGGATTAAGGAGCACTGTGATGATTAAGCTGACAGTGGCCCAAGTTTGAAGTGGT 1611
Qy 1394 ATGCAAGCAGCAAGAAAGCCATCTTGTCAAGAAATTCGTCGAGCGGATAGCTGCA 1453
Db 1612 ATTAAGAAATTAAGAAAGCCATTAATTCAGAGGAGATATGAGACAAAGATTCAGGGT 1671
Qy 1454 TCCACTAGATTCACCTCAATGTTTCTCGAAGAGTACCAAGCAGAGCTCGTTGAAA 1512
Db 1672 TTCAACAGAGATCCACTGTGATGTTCACTGAAAGATACAGAAAGTCTGTAGAGAGT 1731
Qy 1513 ACGATCAGAGCTCTTTGAAAAA-----AGACTACATATCGGAACACAGCTGGGCT 1567
Db 1732 ACCATCTGGGTCTGGATCAAAAACGCAAAATATGTGTGGAGAGCTCATTTTGAAT 1791
Qy 1568 TTGAGATTTTAAAGCTCTCTCAGATGTGAGAAAGCAATTTCTCAACCAAGAGGTGTT 1627
Db 1792 TTGGCGATTTATGATGATGAGAGTCAACGAGAGAGTGTGTGGGAAATTAAGAGGATCT 1851
Qy 1628 TCACAAGACAGACAAACCAACTGTTGCTCATGTACTGAGA 1671
Db 1852 TCACCTGGCAGAGACAAACCAAAAGTGCAGCTTCTTTGCGA 1895

```

RESULT 13

LOCUS RGLUCB 2391 bp mRNA linear ROD 10-FEB-1999
 DEFINITION Rat mRNA for beta-glucuronidase.
 ACCESSION Y00717
 VERSION Y00717.1 GI:56270

KEYWORDS beta-glucuronidase.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 Powell, P.P., Kyle, J.W., Miller, R.D., Pantano, J., Grubb, J.H. and
 Sly, W.S.
 Rat liver beta-glucuronidase. cDNA cloning, sequence comparisons
 and expression of a chimeric protein in COS cells
 Biochem. J. 250 (2), 547-555 (1988)
 88183378

JOURNAL
 MEDLINE
 PUBMED 3355537

REFERENCE
 2 (bases 1 to 2391)
 Miller, R.D.
 Direct Submission
 Submitted (07-JAN-1988) Miller R.D., E.A. Doisy Dept. of
 Biochemistry, St. Louis University School of Medicine, 1402 S.
 Grand Blvd., St. Louis, MO, 63104
 Location/Qualifiers

FEATURES
 source
 1..2391
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="pR165"
 /tissue_type="liver"
 <1..1909
 /codon_start=2
 /product="beta-glucuronidase precursor"
 /protein_id="CA68705.1"
 /db_xref="GI:758260"
 /db_xref="Gene:758260"
 /db_xref="Swiss-Prot:P06760"
 /translation="ELICSCALALOGGLFPEKTPSRELKVLIDGLMSFRADYSNNRLO
 GFEKOWRQPLRESGPTLMDVPSSPNDIQEALRNIGVWREAVLRFQRTQDT
 DRYVLRNSAHYAVVWVNGIHVHEHGGHLPREADITTKAVQSGPLTFRVTLAIN
 TLTPTLPFGTIVYKTDPSMYPKSIPVDISDFEPNAGLRSVVLVTPPTVYIDIT
 VTTPVDVRLVNWVSVQSDHFLQLEVLRLDEBQKVARGTGNGQLKVPRAHLP
 YLMEHPEAYLESLVMTTPESVDFYTLPGIRIVATKSKFLINGKPFYQGVNKH
 ESDIRGPFMDPLIKDFNLRLWGLMGSFRTSHVPSSEVLOCDRGIIVIDECG
 VGIYLPQSGVSLRHLRYNDLVRDKNHNPAYVMSVNAEPYSLSKPAQVYFETLI
 AHTRALDPTFRVTVSNTRDADIGAPRVYICNYSVTHYDGHLEVTQLQTSOF
 ENWRYMTQKPLIQSEYADAVSGAHEDPRAFSESYQTALLENHLLIDERKEVIG
 ELINWFADPMTNOSPRLRVYTNKKGIFTRQNPKNAAFLIRBYRIANETRGYSVPR
 TQCGSRPFPT"
 <1..28
 sig_peptide
 29..1396
 mat_peptide
 /product="beta-glucuronidase"
 2391

ORIGIN
 polyA_site

Query Match 9.9%; Score 166.8; DB 10; Length 2391;
 Best Local Similarity 51.3%; Pred. No. 1.6e-33;
 Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;

```

Qy 751 GACGAGTACTCTGACATCGAATCAGACGATCAGCTGGAGCAGAGAGGCTCTAT 810
Db 902 GACCTCTATACCTCCCTCGTGGGATTCGAACAGTGCCTGTCAAAAGAGCAAGTTCTC 961
Qy 811 CTGAACGGGAAACCTGTTTCTTTTGAAGGCTTTGGAAGCAGAGAAATCCCGCTTG 870
Db 962 ATTAATGGGAAGCCCTTCTTACTCCAGCGCTCAACAGAGATGAGATTCAATTCGA 1021
Qy 871 GGGAGGCGACCTTTTATCATGATGATTAAGACTTCAACCTTCTGAAGTGAATCAAC 930
Db 1022 GGGAGGCTTCTGAGTGCCTCTGTGTATTAAGATTTTCAACCTTCTCGTTGGCTGG 1081
Qy 931 GCGAATTTTCAAGACCTTCACTATCTTACAGTGAAGAGTGGCTGATCTTGGCCGAC 990
Db 1082 GCAATTCCTTTGATACAGCAGCATATCTCTACGAGGAGGATGATCTTCACTGTGAC 1141

```

QY 991 AGACTGGAACTCTTGATAGAGAAAGCCCGACGTTGGTATCAACAAGTACCA---C 1047
 Db 1142 CGATATGAAATGGGTGATCATGATGAGTCCCGGTGTGGGATCGGCTGCCAGAGT 1201
 QY 1048 TACAATCCGAGACTCGAAGATAGCAGAAAGACAACTAAGAAAGATATGACAGAC 1107
 Db 1202 TTGGCAACGATGCTCTTCCGACCACTTGAAGGTGATGACAGAGCTGTGGCCAGAGAC 1261
 QY 1108 AAGAACATCCAGTGTGATCATGTGAGTGTGGCAACAGACAGATCCAACTCA 1167
 Db 1262 AAAAATACCTCGGTGCTGTGATGTGTCTGTGGCCATAGAGCTGTCTTCTTGAAA 1321
 QY 1168 GACGCGAGGGTTCTTCAAGCCCTTATGAGATCCCAATGAAATGATGCAACAGC 1227
 Db 1322 CTGGCCGATATTACTTCAAGACGCTGATGCGCCACCAAGCCCTGGACCCACCCGT 1381
 QY 1228 CCGCTGTCAATGAGACATGATGAGATGAGACGACCAAGAGAAACAGAGAGTGGGCTG 1287
 Db 1382 CCGTGAACCTTTGGAGCAATA-----CGAATATGACGACAGCTGGGGGCC 1429
 QY 1288 AAGTACTGACATCGTCTGTGTGACAGGTAACGCTGTGTAATCTATCAAGGAAAG 1347
 Db 1430 CCGTACGTGACGATGATTTGTGTGAACAGTTACTTATCTGTATCATGACTAGGGCAT 1489
 QY 1348 ATGAGAGAGAGACTTCAAGCTCTGGAAGAAAGACATAGAAAGCTCTATGCAAGGACAG 1407
 Db 1490 CTGAGAGTGAATTCAGCTGACGCTGACTAGCCAGTTTGAAGAACTGTATAGATGTACAG 1549
 QY 1408 AAGCCCATCTTTTGCACAGAAATGATGCGGAGCGAGTAGCTGACATCAAGATCCA 1467
 Db 1550 AAGCAATTTATTCAGAGCGAGATATGAGACAGACCGCTCTCGGGCTTCAAGAGATCCA 1609
 QY 1468 CCTCAATGTTCTCCGAGAGTACCAAGACAGAGCTGTGAAAAGACGATCAAGCTCTT 1527
 Db 1510 CCGTGAATGTTCACTGAGAGATACCAAGACAGCTCTCTGAGATTTATCAATTTGATCTTG 1569
 QY 1528 TTGAA-----AAAAGTACATCATGGAACACAGTGTGGGCTTTGCAAGATTTTAAG 1581
 Db 1670 GATGAGAAACGAAAGAAATATGTCATCGAGAGCTCATCTGGAATTTTGTGACTTATG 1729
 QY 1582 ACTCTGAGAAATGAGAAAGCCATTTCTCAACCAAGAGGTGTTTCAACAAGAGACAGA 1641
 Db 1730 ACGAACCAATGACCACTGAGAGTAAACAGAAACAAAGAGGAGATCTTCACTGACAGAGA 1789
 QY 1642 CAACCCAAATCTGTTGCTCATGTAAGTGA 1671
 Db 1790 AACCCAAAGATGGCAGCTTCATTTGGCA 1819

RESULT 14
 AX827491 2472 bp DNA linear PAT 12-DEC-2003
 LOCUS Sequence 225 from Patent EP1344834.
 DEFINITION AX827491
 ACCESSION AX827491 GI:39837679
 VERSION AX827491.1 GI:39837679
 KEYWORDS
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 Boeser, F., Suter-Dick, L. and Wolf, D.
 AUTHORS Methods for the toxicity prediction of a compound
 JOURNAL Patent: EP 1344834-A 225 17-SEP-2003;
 F. HOFMANN-LA ROCHE AG (CH)
 FEATURES
 1 Location/Qualifiers
 1..2472
 /organism="Rattus norvegicus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:10116"

ORIGIN

Query Match 9.9%; Score 166.8; DB 6; Length 2472;
 Best Local Similarity 51.3%; Pred. No. 1.6e-33;
 Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;

QY 751 GACGATACACTCTGGAACATCGAATCCAGATCAGCTGAGAGAGAGGCTCTAT 810
 Db 946 GACTTCTATACCTCCCTGCGGATTCGACAGTGGCTGTCAAAAGCAAGTTCTTC 1005
 QY 811 CTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAAGCAGAGAAATCCCGTCTTG 870
 Db 1006 ATAAATGGGAACCTTTTCACTTCCAGAGGCTCAACAGCATAGAGATTCAGATATCCGA 1065
 QY 871 GGGCAGGACCTTTTATCATGATGATTAAGAAAGCTTCAACTCTGAAGTGAATCAAC 930
 Db 1066 GGGAGAGGTTGACCTGCTGCTGTGATTAAGATTTTAACTCTCCGTTGGCTCGGG 1125
 QY 931 GCGAATCTTTGAGACCTCTGACTATCCCTTACAGTGAAGATGGCTGATCTTGGCAG 990
 Db 1126 GCAATTCCTTTGTATCCAGCACTATCCCTACTGAGAGAGTACTTCAAGCTGTGAC 1185
 QY 991 AAGCTCGAATCTGTGTATGACAGAGCCCGCACGTTGTATCAACAAGTACCA---C 1047
 Db 1186 CGATATGAAATTTGTGTATCATGATGATGTCTCCGTTGGGATCGTGTGCCCAAGAT 1245
 QY 1048 TACAATCCGAGACTCGAAGATGAGCAAGAACAACTAAGAAATGATGACAGACAC 1107
 Db 1246 TTGGCAAGGTGCTCTTTGGGACCACTTAGAGGTATGAGAGAGTGGTGGCGAGGAC 1305
 QY 1108 AAGAACATCCAGTGTGATCATGTGAGTGTGGGAGAGAAACAGAGTCCAACTCA 1167
 Db 1306 AAAAATCACCTCGGCTGTGATGTGTGTGTGTCGACCAATGAGAGCTGTCTTCTTGAAA 1365
 QY 1168 GACGCGAGGGTTCTTCAAGCCCTTATGAGATGCTCCAAAGAAATGATGAAACAGC 1227
 Db 1366 CCGTCCGATATTTACTTCAAGACGCTAGTCCCAACCAAGCCCTGACCCACCCGT 1425
 QY 1228 CCGTGTGATGATGATGATGATGAGCGACCAAGACAGAAACAGAGAGTGGGCTG 1287
 Db 1426 CCGTGAACCTTTGTGACATA-----CGAATATGACCAAGATGAGGAGG 1473
 QY 1288 AAGTACTGACATCGTCTGTGTGAAACAGTACTACGCTGTGTAATCTATCAAGGAAAG 1347
 Db 1474 CCGTACGTGAGAGTATTTGTGTGAACAGTTACTTATCTGTGATCATGACTAGGGCAT 1533
 QY 1348 ATGAGAAAGAGACTTCAAGCTCTGGAAGAAAGCATGAAAGACTCATAGAGGACAG 1407
 Db 1534 CTGAGAGTGAATTCAGCTGACGCTAGCTACAGTTTGAAGACTGTATTAAGATGTACAG 1593
 QY 1408 AAGCCCATCTTTGTGACAGAAATCGGTGCGAGCGCATAGTGGCATCACTACGATCCA 1467
 Db 1594 AAGCCATTTATTCAGAGCGAGTATGAGACAGACCGCTCTCGGGGCTTCAAGAGATCCA 1653
 QY 1468 CCTCAATGTTCTCCGAGAGTACCAAGACAGAGCTGTGAAAAGACGATCAGGCTCTT 1527
 Db 1654 CCGCATGTTTCACTGAGAGGTACAGACAGCTCTCTGAGAAATTAATCATTTGATCTTG 1713
 QY 1528 TTGAA-----AAAAGTACATCATCGGAACCAAGAGGTGTTTCAACAAGAGACAGA 1581
 Db 1714 GATGAGAAACGAAAGAAATATGTCATCGAGAGCTCATCTGGAATTTTGTGACTTCATG 1773
 QY 1582 ACTCTGAGAAATGAGAAAGCCATTTCTCAACCAAGAGGTGTTTCAACAAGAGACAGA 1641
 Db 1774 ACGAACCAATGACCACTGAGAGTAAACAGAAACAAAGAGGAGATCTTCACTGACAGAGA 1833
 QY 1642 CAACCCAAATCTGTTGCTCATGTAAGTGA 1671
 Db 1834 AACCCAAAGATGGCAGCTTCATTTGGCA 1863

RESULT 15
 RATGUCB 2472 bp mRNA linear ROD 27-APR-1993
 LOCUS Rat beta-glucuronidase mRNA, complete cds.
 DEFINITION

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2005, 07:30:35 ; Search time 25 Seconds
(without alignment)

2166.802 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001
Sequence: 1 MVRPQRMKKRFLILNGVWN.....TRDQPKVAVLRLNSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2899	99.9	563	A72300	beta-glucuronidase
2	1011	33.7	570	C90485	beta-glucuronidase
3	899.5	30.0	648	A32576	beta-glucuronidase
4	894.5	29.8	648	A25047	beta-glucuronidase
5	879	29.3	603	GBRCGC	beta-glucuronidase
6	865.5	28.8	651	A26581	beta-glucuronidase
7	691	23.0	368	A85768	partial beta-D-glu
8	691	23.0	370	D90919	beta-D-glucuronida
9	487	16.2	755	D95842	probable beta-gala
10	482	16.1	716	JU0275	beta-galactosidase
11	481.5	16.0	996	D86872	beta-galactosidase
12	452.5	15.1	1087	F72283	beta-galactosidase
13	414	13.8	1024	1 GBRC	beta-galactosidase
14	405.5	13.5	1024	E90678	beta-D-galactosida
15	405.5	13.5	1024	A85529	beta-D-galactosida
16	381.5	12.7	1014	C83990	beta-galactosidase
17	371.5	12.4	1307	T35944	beta-galactosidase
18	364	12.1	897	A39405	probable beta-gala
19	360.5	12.0	1060	A10201	beta-galactosidase
20	352.5	11.7	1025	JC1266	beta-galactosidase
21	352.5	11.7	1042	1 GBRC	beta-galactosidase
22	352.5	11.7	1042	2 B85968	evolved beta-D-gal
23	352.5	11.7	1042	2 F91123	evolved beta-D-gal
24	348.5	11.6	1034	2 T30574	beta-galactosidase
25	348	11.6	1026	A49750	beta-galactosidase
26	339.5	11.3	1034	2 A24925	beta-galactosidase
27	328	10.9	1075	2 T47603	beta-galactosidase
28	327.5	10.9	1034	2 T30551	beta-galactosidase
29	322.5	10.7	626	2 A42891	beta-galactosidase

30	321	10.7	2228	2 B97942	beta-galactosidase
31	321	10.7	2233	2 B95075	beta-galactosidase
32	312	10.4	1015	2 T39697	beta-galactosidase
33	310	10.3	1005	2 T31333	beta-galactosidase
34	300.5	10.0	1007	2 A30093	beta-galactosidase
35	256	8.5	785	2 H72228	hypothetical prote
36	194	6.5	237	2 B90919	interrupted beta-D
37	194	6.5	237	2 B85768	partial beta-D-glu
38	191.5	6.4	827	2 AB2764	beta-mannosidase p
39	191.5	6.4	832	2 A97545	hypothetical prote
40	181	6.0	820	2 T37230	probable beta-mann
41	179.5	6.0	900	2 T19689	hypothetical prote
42	179	6.0	879	2 A55881	beta-mannosidase (
43	175	5.8	891	2 A82755	beta-mannosidase p
44	139.5	4.6	1091	2 F83928	hypothetical prote
45	132	4.4	818	2 A97668	mannosidase Agr_C

ALIGNMENTS

RESULT 1					
A72300 beta-glucuronidase - Thermotoga maritima (strain MSB8)					
C.Species: Thermotoga maritima					
C.Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004					
C.Accession: A72300					
R.Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey					
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;					
C.M.					
Nature 399, 323-329, 1999					
A.Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq					
A.Reference number: A7200; PMID:99287316; PMID:10360571					
A.Accession: A72300					
A.Status: preliminary					
A.Molecule type: DNA					
A.Residues: 1-563 <ARN>					
A.Cross-references: UNIPROT:Q9X0F2; GB:AE001766; GB:AE000512; NID:G4981600; PIDD:AMD3614					
A.Experimental source: strain MSB8					
C.Genetics:					
A.Gene: TM1062					
C.Superfamily: beta-glucuronidase					
Query Match					
Best Local Similarity 99.9%; Score 2899; DB 2; Length 563;					
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	1	MVRPQRMKKRFLILNGVWNLEVTSKDRPIAVPGSMNBOYODLCYEEGPTTKTFYVPK	60		
DB	1	MVRPQRMKKRFLILNGVWNLEVTSKDRPIAVPGSMNBOYODLCYEEGPTTKTFYVPK	60		
QY	61	XLISQGHILYFAAVNTDCEVFLNGEKVGENHIEYLPEVDVYGVKSGENELRVVENRL	120		
DB	61	ELISQGHILYFAAVNTDCEVFLNGEKVGENHIEYLPEVDVYGVKSGENELRVVENRL	120		
QY	121	KVGGPPSKVPSGHTTVPFGSGFPFANDFPPYGGIIRPVLIIEFDHARILDIWDTSES	180		
DB	121	KVGGPPSKVPSGHTTVPFGSGFPFANDFPPYGGIIRPVLIIEFDHARILDIWDTSES	180		
QY	181	EPEKLGKVVKKIVSEEAQVQEMTKLGESEKRTSRNRFEGEFTLENNRFLSDPY	240		
DB	181	EPEKLGKVVKKIVSEEAQVQEMTKLGESEKRTSRNRFEGEFTLENNRFLSDPY	240		
QY	241	LYPLKVELKEDYTLIDIGIRITSMDEKRLYNGKVFPLKGFQKHEEPVLGGTFYPLMI	300		
DB	241	LYPLKVELKEDYTLIDIGIRITSMDEKRLYNGKVFPLKGFQKHEEPVLGGTFYPLMI	300		
QY	301	KDFNLKVINANSFRTSHYPISEEWLADRLGILVIDEAPHVGTTRHYNPETOKIAD	360		
DB	301	KDFNLKVINANSFRTSHYPISEEWLADRLGILVIDEAPHVGTTRHYNPETOKIAD	360		
QY	361	NIRRMIDHKNHPSVIVMSVANEPSNHPDAGEFKALYETANEMDRRPVVMVSMMDAP	420		
DB	361	NIRRMIDHKNHPSVIVMSVANEPSNHPDAGEFKALYETANEMDRRPVVMVSMMDAP	420		

Db 361 NIRMIDRKHNSVIMSVANEPESNHPDAGFFKALYETANEMDRTPVVMVSMADP 420

Qy 421 DERTDVALKYPDYICVNRYYGYIYQRIEGLQALEKDIIELYARHKPIFTVEGAD 480

Db 422 DERTDVALKYPDYICVNRYYGYIYQRIEGLQALEKDIIELYARHKPIFTVEGAD 480

Qy 481 AINGHYDPDPQMFSEYQALVEKTRILLLKKDYIIGTHWAFADPKTPONVRRPILNHK 540

Db 481 AINGHYDPDPQMFSEYQALVEKTRILLLKKDYIIGTHWAFADPKTPONVRRPILNHK 540

Qy 541 GVFTDRDPKLVAVHLRLMSEV 563

Db 541 GVFTDRDPKLVAVHLRLMSEV 563

RESULT 2

beta-glucuronidase (gnus) [imported] - Sulfolobus solfataricus

C/Species: Sulfolobus solfataricus

C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C/Accession: C90485

R/She: Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Amayez, M.J.; Chan-Jong, I.; Jeffries, A.C.; Kozera, C.U.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F. submitted to GenBank, April 2001

A/Description: Sulfolobus solfataricus complete genome.

A/Reference number: A99139

A/Accession: C90485

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-570 <KDN>

A/Cross-references: UNIPROT:Q97UI1; GB:AE006641; NID:g13816434; PIDN:AAK4138.1; GSPDB:CA:Genetics: gnus

C/Superfamily: beta-glucuronidase

Query Match 33.7%; Score 1011; DB 2; Length 570;

Best Local Similarity 39.2%; Pred. No. 2.7e-61;

Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

Qy 15 LNWVNNLEVTSKDR-----IAPGSSNNEQYODLYEBCPFYKTFYVPK 60

Db 11 LQGFNFKIDENNGEENGWYKGLSESDIIVPASMENQPKMDQFGIMYQKDLVSN 70

Qy 61 XLGKHRLVPAVNTDCEVFLNGKGVENHIEVLPEVDVTGKVGSGENELRVVENRL 120

Db 71 DNGRKAAMVPEGAGYITKMLINEYGTGHTGCTQKFKPKLVY---NEFKIV---V 123

Qy 121 KVGGFPSKVPDSTGHTVGFEGSPPPAN-----DFPPYGGIIRPVLIETDHAIRLIDW 174

Db 124 KIDWTPSPY-----NLPPARDLNNAAPDFNYGGIHRPVYIEFVDECHVEDIT 171

Qy 175 VDTSESEBEKLLGVKVKIYSEEAQVQEMTIGSEBEKI---KISNRFVGEFLEN 230

Db 172 VYI-----KSYGHLKVEI-LSECNQRFSLFKLVDEGRVILNEESSNVEFXD--VNN 222

Qy 231 ARFWSLDDPYLYPKVLE-----KDEYTLIDIGITISMEKRLYLNGKFEVLPQKHE 285

Db 223 VIPSPPNPVLYTLIVEMVYGNKADSVIRKIGRDVENVQDKIYLNGKIFLKGFRHE 282

Qy 286 EFPVLGGCTPYPLMKIPNLKIMANSTSHYPSYSEWMLDLADRLGILVIDEAP--HV 343

Db 283 DFPILGFTYGAVALVRDFFLMRKIGANSFRTSHYPSENHDLADEWGFVILIEPLCYG 342

Qy 344 GITFVHNPEYOK-----AEDNIRIMDRKHNSVIMSVANEPESNHPDAEGF 394

Db 343 NISVYMQEELANFGVYKFEKVRDTIKEMIRQHKRPSVIVYSVNPEPSDIREVAEP 402

Qy 395 FKALYETANEMDRTPVVMVSMADPDERDVALKYPDYICVNRYYGYIYQRIEGL 454

Db 403 IRREVELFKSLDSRPVTFAS-----HRSVRDLALELVVDVVISLNYHGWTFEGDDISGV 457

Qy 455 QALEKDIIELYARH-RKPIFTVEGADALINGHYDPDPQMFSEYQALVEKTRILLLKKD 513

Db 458 KVAIELEELHKKPEKPIITTEFGADALYGLHSDPQWMSSEYQSMIRKYIEALREKD 517

Qy 514 YIIGTHWAFADPKTPONVRRPILNHKGVFTDRDPKLVAVHLRLMSEV 560

Db 518 YIIGTHWAFADPKTPONVRRPILNHKGVFTDRDPKLVAVHLRLMSEV 564

RESULT 3

beta-glucuronidase (EC 3.2.1.31) allele B precursor - mouse

N/Alternate names: beta-D-glucuronoside glucuronosylhydrolase

C/Species: Mus musculus (house mouse)

C/Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 09-Jul-2004

C/Accession: A32576; B32576; I49692; A28954; A29977; A35798

R/Mawzyntak, C.U.; Gallagher, P.M.; D'Amore, M.A.; Carter, J.E.; Lund, S.D.; Rinchik, E. Mol. Cell. Biol. 9, 4074-4078, 1989

A/Title: DNA determinants of structural and regulatory variation within the murine beta-g

A/Reference number: A32576; MUID:89384641; PMID:2779578

A/Accession: A32576

A/Molecule type: mRNA

A/Residues: 1-648 <NAM>

A/Cross-references: UNIPROT:P12265; GB:M28540; GB:M27816; NID:g193718; PIDN:AAA63307.1; I

A/Experimental source: allele B

A/Accession: B32576

A/Molecule type: mRNA

A/Residues: 1-86; 'T', 88-648 <NA2>

A/Cross-references: GB:M28541; NID:g193720; PIDN:AAA63308.1; PID:g193721; GB:M27816

A/Experimental source: allele H

R/Funkstein, B.; Leary, S.L.; Stein, J.C.; Caterall, J.F. Mol. Cell. Biol. 8, 1160-1168, 1988

A/Title: Genomic organization and sequence of the Gue-B-a allele of the murine beta-gluc

A/Reference number: I49692; MUID:88216590; PMID:2835664

A/Accession: I49692

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-232; 'T', 234-264; 'D', 266-319; 'V', 321-427; 'K', 429-615; 'L', 617-648 <RES>

A/Cross-references: GB:M19279; NID:g193524; PIDN:AAA37697.1; PID:g309257

A/Experimental source: allele A

R/D'Amore, M.A.; Gallagher, P.M.; Korfhagen, T.R.; Ganschow, R.E. Biochemistry 27, 7131-7140, 1988

A/Title: Complete sequence and organization of the murine beta-glucuronidase gene.

A/Accession: A28954; MUID:89062453; PMID:3196706

A/Molecule type: DNA

A/Residues: 1-264; 'D', 266-319; 'V', 321-648 <DAN>

A/Cross-references: GB:J02836; GB:J03035; GB:M20204; NID:g193716; PIDN:AAA98623.1; PID:g

R/Gallagher, P.M.; D'Amore, M.A.; Lund, S.D.; Ganschow, R.E. Genomics 2, 215-219, 1988

A/Title: The complete nucleotide sequence of murine beta-glucuronidase mRNA and its deduc

A/Reference number: A29977; MUID:88284700; PMID:3397060

A/Accession: A29977

A/Molecule type: mRNA

A/Residues: 1-264; 'D', 266-319; 'V', 321-648 <GAL>

A/Cross-references: GB:J03047; NID:g193522; PIDN:AAA37696.1; PID:g309256

R/Ii, H.; Takeuchi, K.H.; Manly, K.; Chapman, V.; Swank, R.T. J. Biol. Chem. 265, 14732-14735, 1990

A/Title: The propeptide of beta-glucuronidase. Further evidence of its involvement in cor

A/Reference number: A35798; MUID:90368633; PMID:2394691

A/Accession: A35798

A/Status: not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 593-648 <LIA>

A/Note: the location of the propeptide cleavage site was not demonstrated directly but is

C/Comment: in some tissues, a portion of this enzyme is retained in the endoplasmic retic

C/Genetics:

A:Gene: Gns

A:Map position: 5

A:Functions: 70/3; 132/3; 193/2; 241/1; 303/3; 351/3; 411/2; 460/2; 488/3; 547/3; 593/1

C/Superfamily: beta-glucuronidase

C/Keywords: glycosidase; hydrolase; lysosome

F:1-22/Domains: signal sequence #status predicted <SIG>
 F:23-648/Product: beta-glucuronidase, ER-retained form #status predicted <ERMT>
 F:23-643/Product: beta-glucuronidase, lysosomal (default) form #status predicted <LMT>
 F:63-648/Domains: carboxyl-terminal propeptide #status predicted <CPRO>

Query Match 30.0%; Score 899.5; DB 2; Length 648;

Best Local Similarity 35.9%; Pred. No. 1.3e-53;

Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;

```

Qy 1 MVRPQRKKRFTLLINGVWNLGV-TSKDR-----PIAVGSMNEQ 39
Db 27 MLFKESPSRELKALDGLMFRADLSNNRLOGFEQWYQPLRESGPTLDMPVPSFNDI 86
Qy 40 YQDLCYEE--GPFYKTFYVKKLSQ--KHRLRYAANTDCEVFLNGEKVGNHLEY 94
Db 87 TQEAALRFQGWWMYERBAVLPQRKTQDTDRVLRINSAAHYAVVWNGIHHVEHEGSH 146
Qy 95 LPFEVDVGVKYSQ---ENELRVVENRLKYGFPSPKVP-----DSGTHVGFSGSF 143
Db 147 LPFEADITKLVQSGPLTFRVTIAINNTLT----PYTLPGFTIYKTDPSMYPKGYF--V 200
Qy 144 PRANPDPFPYGGIIRPVLIETFDHARILDIWDTSSESEPEKKLGKVK--VKIEVSEEA 201
Db 201 QDTSFDFPNYAGLHRSVLYTTPTYIDITVITN---VEODIGLVYMWISVQSGEHF-- 255
Qy 202 QEMTIKLGEEKKIRTSNRPVEGEFLENNARFV----SLEDP-YLYPLKYV----ELEKD 251
Db 256 -QLFVQLDDEGGKVAHGTGQGLQVPSANLWMPYLMHEHPAVWYSLFVVTTESVTD 314
Qy 252 EYTLIDIGRTISWDEKRLYLNGKRPVFLKGFEKHEEPVLQGTFFYPLMIKDIFNLKWINA 311
Db 315 YTLPIGRTIYVATKSKFLNGKRPYFGQVNGKHEDSDIRGKGFMDPLVAKDFNLKWLGA 374
Qy 312 NSPFTSHYPSYSEEMLDLADRLGLIVIDEAPHVGIT-----RYHNPETOKIAEDN 361
Db 375 NSPFTSHYPSYSEEVYQLCDRYGIVVIDECPGVGILPQSPGNESLRHNL-----EWMEL 429
Qy 362 IRRNIDRKHKRPSTVIMSVANEPESNHPDAGFFGALYETANEMDRTPPVWVS--MMDA 419
Db 430 VR----RDKNHPAVWMSVANEPSSALKPAAVYFKTLITHTKALDLTPVTFVSNAKYDA 485
Qy 420 PDERTRDVALKCFDVCNRYGYWYIYQGRIEBGLQALEKDIIEELVYARHKRPIFTEGGA 479
Db 486 -----DLGARYVDVICNNSYFSWHDYGHLEVIQPLNSQPENNYKTHQKPIIOSEGA 539
Qy 480 DALAGIHYPQPMSESEYQAEIVLVEKTRILL--LKDYIIGTHWAFADFKTPQNVREPI 537
Db 540 DALGHEHDPKPMSESEYQKAVLENYHSLVDQKREYVAGELINMFADFMNQSPRLAVIG 599
Qy 538 NHKGVFTDRQPKLVAVHVR-RLW 560
Db 600 NKGIFTRQRPKTSAPILRERYW 623

```

RESULT 4

A:Accession: A25047

A:Residues: 1-648 <NTS>

A:Cross-references: UNIPROT:P06760; GB:M3962; NID:g204329; PID:AAA1228.1; PID:g204330

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: A25047

A:Residues: 1-648 <NTS>

A:Cross-references: UNIPROT:P06760; GB:M3962; NID:g204329; PID:AAA1228.1; PID:g204330

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Accession: S00345

A:Molecule type: mRNA

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: S00345

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: S00345

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: S00345

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: S00345

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: S00345

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: S00345

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Accession: S00345

A:Residues: 1-648 <NTS>

A:Cross-references: EMBL:Y00717; NID:g56270; PID:CAA68705.1; PID:g758260

A:Experimental source: female preputial gland

R:Powell, P.P.; Kyle, J.W.; Miller, R.D.; Pantano, J.; Grubb, J.H.; Sly, W.S.

Biochem. J. 250, 547-555, 1988

A:Title: Rat liver beta-glucuronidase, cDNA cloning, sequence comparisons and expression

A:Reference number: S00345; MUID:86183378; PMID:3335537

A:Molecule type: mRNA

A:Residues: 1-603 <BLAT>
 A:Cross-references: UNIPROT:P05804; GB:A5000257; GB:U00096; NID:g1787898; PIDN:AACT4689.
 A:Experimental source: strain K-12, substrain M61655
 R:Schlaman, H.R.; Risseuw, E.; Franke-van Dijk, M.E.; Hooykaas, P.J.
 Gene 138, 259-260, 1994
 A:Title: Nucleotide sequence corrections of the uida open reading frame encoding beta-g1
 A:Reference number: 153717; MUID:94171050; PMID:8125312
 A:Accession: 153717
 A:Status: translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-603 <RES>
 A:Cross-references: GB:S69414; NID:9545893; PIDN:AA30197.1; PID:9545894
 R:Jefferson, R.A.; Burgess, S.M.; Hirsch, D.
 Proc. Natl. Acad. Sci. U.S.A. 83, 8447-8451, 1986
 A:Title: Beta-glucuronidase from *Escherichia coli* as a gene-fusion marker.
 A:Reference number: A26487; MUID:87041472; PMID:3534890
 A:Accession: A26487
 A:Molecule type: DNA
 A:Residues: 1-419; 'VHGNIS', 427-603 <JEF>
 A:Cross-references: GB:M14641; NID:g868017; PIDN:AAA68923.1; PID:g868020
 R:Punt, P.
 Submitted to the EMBL Data Library, April 1994
 A:Reference number: S43555
 A:Accession: S43555
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-'V', 3-603 <PUN>
 A:Cross-references: EMBL:Z37701; NID:g475168; PID:g475169
 C:Genetic: This acid hydrolase catalyzes the cleavage of a wide variety of beta-glucuron
 A:Gene: uida
 A:Map position: 36 min
 C:Function:
 A:Description: catalyzes hydrolysis of beta-D-glucuronoside to D-glucuronate and alcohol
 A:Pathway: strach and glucose metabolism
 C:Superfamily: beta-glucuronidase
 C:Keywords: glycosidase; hydrolase

Query Match 29.3%; Score 879; DB 1; Length 603;
 Best Local Similarity 33.8%; Pred. No. 3e-52;
 Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

```

QY 1 MVRPQNRKKRFFIILNGVNNLEV-----TSKDRPIAVPGSMNEQYDLCY 45
DB 1 MVRPVRPTREIKKLDLMAFSIDRENGCIDQRMWESALGSRALAVPGSFNDQFADADI 60
QY 46 EE--GPTTYKTFYVPPKLSQKHIRLYPAVNTDCEVFLNGKXGHNHIEYLPREVDTG 103
DB 61 RNVAQNVWYOREVAFIPKMGAGORIVLRFDAVTHGKAVNNQEVMEHQGGYTFPEADVTP 120
QY 104 KVSNGEN-ELRVVVENLTKVGGFPSKV---PDGTHVVGFGSPFPPANPDPFGGIRP 159
DB 121 YVLAGKSVRTVCNNELNMOTTPPGNVITDENCKKKQSYF-----HDFPNYAGIHRS 173
QY 160 VLEFTHARILIDVDTSESEPEKKLGKVKVLESEAVGQEMTIKLGEEKKRTSN 219
DB 174 VMLYTTNTVWDITVTTHVAQ---DCNHSVDMQV---VANGVSVSLADADQOVATG 227
QY 220 RVEGEFLENAFWSLDEPPLYL---KVELKDEYTLIDIGIRTSIMDEKRLYLNGKP 275
DB 228 QGTSIGTLQVNVPHLMQGEGLYELCVTAKSQTECDIYPLRVGIRSAVAVGEGCLINHKP 287
QY 276 VFLKSGKHEEFPVLGGTFPPLMIKDPNLLKXMINASPTSHVPESEMLDLADRIGIL 335
DB 288 FYFTGFRHEDADLRGKGFNDVNLVHDALMDWIGANSYRFSHPVAEEMLDVADDEHIV 347
QY 336 VIDEAPVVGI-----TRYHNPETOKLAEDNIRIMRHKXHPVIM 377
DB 348 VIDETAIVGPNLSIGIFEGAKNKKEYSEAVNGEFOQAHLOIKRLIARDKNHPEVVM 407
QY 378 MSVANEPESHNPDAEGFPKALYETANEMDRTPVVMVSM--DAPDERTDVALKYFDIV 435
DB 408 WSIANEEDTRPGQAREYFAFLAETRTKLDPTRPITTCVNVVFCDAHTDTISDL-----EDVL 463

```

QY 436 CVNRVYGYTYOGRIEEGIQALEKDIIELYARHRKPIFTEFGADALAGIHVDPQMFSE 495
 DB 464 CLNRYYGVQSSDLETAERVELKEKLMAOEKKHQPIITTEYGVDTLAGHSWYTDWMS 523
 QY 496 EYQAEIVERTIRLLKDDYIIGTHWAFADFKTPQVNRBPLLNHKGVFTRDRPKLVAVH 555
 DB 524 EYQCAWLMDWYHRFDRVSAVVGQVNNFADFNISQGLINVGKKGIFTRDRPKSAFL 583
 QY 556 LRRIMS 561
 DB 584 LQKRW 589

RESULT 6
 A26581
 beta-glucuronidase (EC 3.2.1.31) precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 05-Oct-1988 #sequence revision 05-Oct-1988 #text_change 09-Jul-2004
 C:Accession: A26581; A40337; A24983; A36538
 R:Ohshima, A.; Kyle, J.W.; Miller, R.D.; Hoffmann, J.W.; Powell, P.P.; Grubb, J.H.; Sly, W.
 Proc. Natl. Acad. Sci. U.S.A. 84, 685-689, 1987
 A:Title: Cloning, sequencing, and expression of cDNA for human beta-glucuronidase.
 A:Reference number: A26581; MUID:87118233; PMID:3468507
 A:Accession: A26581
 A:Molecule type: mRNA
 A:Residues: 1-651 <OSH>
 A:Cross-references: UNIPROT:P08236; GB:M15182; NID:g183222; PIDN:AAA52561.1; PID:g183223
 A:Experimental source: Placenta
 R:Shipley, J.M.; Miller, R.D.; Wu, B.M.; Grubb, J.H.; Christensen, S.G.; Kyle, J.W.; Sly, W.
 Genomics 10, 1009-1018, 1991
 A:Title: Analysis of the 5' flanking region of the human beta-glucuronidase gene.
 A:Reference number: A40337; MUID:92009900; PMID:1916806
 A:Accession: A40337
 A:Molecule type: DNA
 A:Residues: 1-70 <SHI>
 A:Cross-references: GB:M65002; NID:g183706; PIDN:AAA52622.1; PID:g183707
 R:Guise, K.S.; Korneluk, R.G.; Wayer, J.; Lamhonwah, A.M.; Quan, F.; Palmer, R.; Ganaschow, A.
 Gene 34, 105-110, 1985
 A:Reference number: A24983; MUID:85232043; PMID:3924735
 A:Accession: A24983
 A:Molecule type: mRNA
 A:Residues: 520-585 <GUI>
 A:Cross-references: GB:M10618; NID:g183704; PIDN:AAA52621.1; PID:g183705
 R:Tomatsu, S.; Fukuda, S.; Sukegawa, K.; Ikeda, Y.; Yamada, S.; Sasaki, T.; C
 Am. J. Hum. Genet. 48, 89-96, 1991
 A:Title: Mucopolysaccharidosis type VII: characterization of mutations and molecular het
 A:Reference number: A36538; MUID:91090114; PMID:1702266
 A:Accession: A36538
 A:Molecule type: mRNA
 A:Residues: 378-385, 616-621, 643-651 <TOM>
 C:Genetic:
 A:Gene: GDB:GUSB
 A:Cross-references: GDB:120025; OMIM:253220
 A:Map position: 7q22-7q22
 C:Superfamily: beta-glucuronidase
 C:Keywords: glycoprotein; glycosidase; homotetramer; hydrolase; lysosome
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-651/Product: beta-glucuronidase, placental #status predicted <MAT>

Query Match 28.8%; Score 865.5; DB 2; Length 651;
 Best Local Similarity 34.5%; Pred. No. 2.8e-51;
 Matches 215; Conservative 105; Mismatches 217; Indels 87; Gaps 21;

```

QY 1 MVRPQNRKKRFFIILNGVNNLEVTSD-----RPI-----AVPGSMNEQ 39
DB 27 MLYPQSPSRECKELGLMSFRADFSDNRRGFEQWYRPLWESGPTVDMVPSSFNDI 86
QY 40 YQD--LCYEGPPTYKTFYVPPKLSQ---KHIRLFAVNTDCEVFLNGKXGHNHIEY 94
DB 87 SQDWRIRHFGVWYVREYVILPERWTQDTRTRVLTIGSAHSYVAIVWVGVDTLHEGGY 146
QY 95 LPREVDTGKVKSG--ENEIR--VVENRLKVGGFPSKVP-----DSGTHVGFSG 142

```

```
Db      147 LPEFADINLVGVGLPRLRTIAINLT-----PTLPECTIQVLDTSKPGYF-- 200
Qy      143 PPANPDEFPYGGIIRPVLEIETDHAAILDIWDTSSESEPEKLGKVKKIEVSEAVQ 202
Db      201 VQNTFDEFPNVAIGLQSRVLLYTPPTVYIDITVTTSS---VEQDSGLVNVQISVKNSL-F 256
Qy      203 EMTITLGESEKKIRSNRFVEGEFLENNRFW-----SLEDP-YLYPLKVELE----- 249
Db      257 KLEVLLDAENKVVANVTGTOGLKVPVSLVWPYLMERPAVYLSLEVOILAQTSGLGPV 316
Qy      250 KDEYLDIGIRITISDEKRLVNGKRVFLKGFKEEFPVLGQTFYPLMKIDPFLMKVI 309
Db      317 SDFYLPVIGIRVAVATKQQLNGKPFYHGVNKIEDADIRKGDWPLVVDPFLMKVL 376
Qy      310 NANSRTSHYPSSEWMLDADRLGILVIDEAPHVG-----ITRYHNPETOKIAD 360
Db      377 GANAFTSHYPAEEMQCDRGIIVIDECPGVGLAPQFPNNVSLHH-----NQVME 432
Qy      361 NIRMIDHKKHPSVIMSVANEPESNHPDAGFPKALYETANEMDRTRPPVVMGMDAP 420
Db      433 VVR-----RDKHPAVVMMSVANEPASHLESAGYILKVIATKSLDPSRPVTVS---N 484
Qy      421 DERTDVALKTYFDICVNRVYGYMYIQGRIEGLQALEKDIELVARRKPIFVTEFGAD 480
Db      485 SNYADKAPYDVICLANSYSYWHYDGHLEIQQLATQFNNYKTKYOKPIIOSEYGH 544
Qy      481 AIAGIHYPDPOMFSEYQAELEVKTRILL---KKDYIIGTHVMAFADFKTPQNVRRPL 537
Db      545 TIAGHPORPLMFTBEYQKSLLEQ-YHLGLDQKRKYVVGELIWMFADPMTQSPTRVLG 603
Qy      538 NHKGVFTDRQPKLVAVLR-RLW 560
Db      604 NKGIFTRQRPKSAFLRERYW 627
```

RESULT 7

```
A:beta-D-glucuronidase [imported] - Escherichia coli (strain 0157:H7, substrain ED
A:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A85768
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaeser, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85768
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-368 <STO>
A:Cross-references: UNIPROT:Q8X671; GB:A0005174; NID:g12515602; PIDN:AAG56605.1; GSPDB:G
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: uidA_1
```

```
Query Match      23.0%; Score 691; DB 2; Length 368;
Best Local Similarity 39.6%; Pred. No. 1e-39;
Matches 141; Conservative 62; Mismatches 125; Indels 28; Gaps 4;

Qy      230 NARFSLDPLVLP-----KYLEKDEYTLDIGRTISWDEKRLYLNGKRVFLKFGKHE 285
Db      3 NPHLMQPEBGVLYELVYTAKSRTBCDIPLRVGIRSVAVKGFLLNHKPFYTFGGRHE 62
Qy      286 EFPVLGQTFYPLMKIDPFLMKVINANSFRTSHYPSSEWMLDADRLGILVIDEAPHYGI 345
Db      63 DADLRGKFPDVLWVHDALMDWIGANSYRSHYPAEEMDKMADHEGIVVIDEFAAAGF 122
Qy      346 -----TRYHNPETOKIADNIRMIDRHKKHPSVIMSVANEPESN 387
Db      123 NLSIGIGFEAGNKPKELYSEBAVNGETQAHLOAIKELIARDKNHPSVIMSVANEPDTR 182
Qy      388 HPDAGFPKALYETANEMDRTRPPVVMGMM--DADDETRDVALKYPVICVNRYYGYI 445
```

```
Db      163 PQGAREYAPLAETARKLDPRTRPTCVVMFCDATDTITSLD-----FDVLCLNRYYGMV 238
Qy      446 YQRIEGLQALEKDIIEELVARRKPIFVTEFGADAIAGIHYPDPOMFSEYQAEVEXT 505
Db      229 QSGDLETAKVLEKELWMOEKHQPILITEYGVDTLAGLSMTYDMSEEQACWMDMY 298
Qy      506 IRLILKKOYIIGTHVMAFADFKTPQNVRRPILNHKGVFTTRDQPKLVAVLRRLMS 561
Db      229 HRVFDVSAVVGEOVMNFPADPATSGILRVGKNKGIPTDRKPKSAFLLQKRW 354
```

RESULT 8

```
beta-D-glucuronidase [imported] - Escherichia coli (strain 0157:H7, substrain RMD 05099
A:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: D90919
R:Hayaishi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90919
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <HAY>
A:Cross-references: UNIPROT:Q8X671; GB:BA000007; PIDN:BA835747.1; PID:g13361791; GSPDB:G
A:Experimental source: strain 0157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC82324
```

```
Query Match      23.0%; Score 691; DB 2; Length 370;
Best Local Similarity 39.6%; Pred. No. 1e-39;
Matches 141; Conservative 62; Mismatches 125; Indels 28; Gaps 4;

Qy      230 NARFSLDPLVLP-----KYLEKDEYTLDIGRTISWDEKRLYLNGKRVFLKFGKHE 285
Db      5 NPHLMQPEBGVLYELVYTAKSRTBCDIPLRVGIRSVAVKGFLLNHKPFYTFGGRHE 64
Qy      286 EFPVLGQTFYPLMKIDPFLMKVINANSFRTSHYPSSEWMLDADRLGILVIDEAPHYGI 345
Db      65 DADLRGKFPDVLWVHDALMDWIGANSYRSHYPAEEMDKMADHEGIVVIDEFAAAGF 124
Qy      346 -----TRYHNPETOKIADNIRMIDRHKKHPSVIMSVANEPESN 387
Db      125 NLSIGIGFEAGNKPKELYSEBAVNGETQAHLOAIKELIARDKNHPSVIMSVANEPDTR 184
Qy      388 HPDAGFPKALYETANEMDRTRPPVVMGMM--DADDETRDVALKYPVICVNRYYGYI 445
Db      185 PQGAREYAPLAETARKLDPRTRPTCVVMFCDATDTITSLD-----FDVLCLNRYYGMV 240
Qy      446 YQRIEGLQALEKDIIEELVARRKPIFVTEFGADAIAGIHYPDPOMFSEYQAEVEXT 505
Db      241 QSGDLETAKVLEKELWMOEKHQPILITEYGVDTLAGLSMTYDMSEEQACWMDMY 300
Qy      506 IRLILKKOYIIGTHVMAFADFKTPQNVRRPILNHKGVFTTRDQPKLVAVLRRLMS 561
Db      301 HRVFDVSAVVGEOVMNFPADPATSGILRVGKNKGIPTDRKPKSAFLLQKRW 356
```

RESULT 9

```
probable beta-galactosidase (EC 3.2.1.23) [imported] - Sinorhizobium meliloti (strain 10
A:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: D95842
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmeester, J.; Chain, P.; Vorholter, F.J.; Herman
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95842
A:Status: preliminary
```

A:Molecule type: DNA
 A:Residues: 1-755 <KUR>
 A:Cross-references: UNIPROT:Q92AF7, GB:AL591985, PIDN:CA048404.1, PID:g15139876, GSPDB:Q
 A:Experimental source: strain 1021, megaplasmid p5ymb
 R:Galibert, F.; Flahm, T.M.; Long, S.R.; Puhler, A.; Apola, P.; Ampe, F.; Barloy-Hubler,
 Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 Hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: lacZ1, SMD21655
 A:Genome: plasmid
 C:Superfamily: beta-galactosidase
 C:Keywords: glycosidase; hydrolase

Query Match 16.2%; Score 487; DB 2; Length 755;
 Best Local Similarity 25.7%; Pred. No. 2.5e-25;
 Matches 155; Conservative 90; Mismatches 193; Indels 164; Gaps 24;

Qy 28 RPIAVGSMNE-----QYODLCYERGPFTYKTFYVKKLSQKHRLYFAAVNTDCEYFLN 83
 Db 30 QPISLPHNAVELPFNYFDERCYORA-FTYGRVLAAMPDPSGREVSLVFDAMADAVYVLN 88
 Qy 84 GEKXGHNIEVLPREVVDYTKVKGSGENELRVVNNRLKVGSPSKVDSGHTVYGFPGSF 143
 Db 89 GEEIVAHNDGYTPPEARLTDRLEGDMLITVKID-----GSENPETPPFG----- 133
 Qy 144 PPAVDFPFGGIRPVLIEFTDARILDIWDTSE-----SEPE----- 183
 Db 134 --GIDVITVAGIYRDWLKTDTPVSANIKIETRDVLSDTKAVSLKCDLSNPGSLPFG 191
 Qy 184 ----KILGVKVKIEVSEAVGQEMTIKLGEEBKJRTSNRFVEGFTLENNAPWSLED 238
 Db 192 TISALLKAAAGBVLAEVAGETTGOSLAEM-----DG---LKGLSLMDID 234
 Qy 239 PVLPLKVELE---KDEYTLIDIGRTISWDEKRLYLNGKRVFLKSGKEEPVLDQGT 294
 Db 235 PVLVIVIELELTGQSDCPAAHFGFRTAEFTTEGFRNGRLKIRGNRQSPFYGVAM 294
 Qy 295 FYPLMKDFNLKLV-INANSFRTSHYPSSEMLDLADRLGILVIDEAP---HVGITRYHY 350
 Db 295 GRTAQBBDADIMGRILHONLVRTSHYQSKWFLDHCRIGLVLEELIPGMOHIC----- 348
 Qy 351 NPETOKIAEDNIRMDHKNHPSVIMSV-ANEPESNHPDAGSEFFALYETANEMDRTR 409
 Db 349 GEEWKOEARVNRMIERDMNHPISIVIGVAINESODSH---DFVATRLARELDPTR 404
 Qy 410 PVVWVSMMDADERTRDVALKCFDVCNRYGYIYQGRLEBGLQLEKDI---BELVA 466
 Db 405 -----QTGVR-----YITDSELELVYTMDFILGNEBPLG 436
 Qy 467 RHRR-----PIFVTEFGADALAGIHYDPMPMSEEV-QAEIVERTIRL 508
 Db 437 ANRBRTRALRPOEGCTGLPRKVPYLITEFGG-----HMVPTKIYDQGRQAEHYRRLLE- 489
 Qy 509 LKKDY----LIGHVAFADFKTPQNVRRPILNHK-----GVTRDRROP 553
 Db 490 VLNAAYDPSGISGAIIGWCMFDYNT-----HKDFGSGDRICYHGVMDFRPF 539
 Qy 554 HV 555
 Db 540 YV 541

RESULT 10
 JU0275
 beta-galactosidase (EC 3.2.1.23) - *Thermoaerobacterium thermosulfurigenes*
 C:Species: *Thermoaerobacterium thermosulfurigenes*
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004

C:Accession: JU0275; JT0963
 R:Burchardt, G.; Bahl, H.
 Gene 106, 13-19, 1991
 A:Title: Cloning and analysis of the beta-galactosidase-encoding gene from *Clostridium tr*
 A:Reference number: JU0275; MUID:92039055; PMID:1840542
 A:Accession: JU0275
 A:Molecule type: DNA
 A:Residues: 1-716 <BUR>
 A:Cross-references: UNIPROT:P26257, GB:M57579, NID:g144838, PIDN:AAA23249.1, PID:g144839
 A:Experimental source: strain EMI
 A:Accession: JT0963
 A:Molecule type: protein
 A:Residues: 1/X', 3-27 <BU2>
 C:Genetics:
 A:Gene: lacZ
 C:Superfamily: beta-galactosidase
 C:Keywords: glycosidase; homodimer; hydrolase
 F:389,429,462/Active site: Glu, Tyr, Glu #status predicted

Query Match 16.1%; Score 482; DB 2; Length 716;
 Best Local Similarity 27.3%; Pred. No. 5e-25;
 Matches 166; Conservative 91; Mismatches 201; Indels 150; Gaps 27;

Qy 30 IAVGSMNEQYODLCYEEG-----PFTY-----KTFYVP 59
 Db 5 IPINNMW---YFKADYEEGKVDLRSFENNVLPHNTIELPYNVDEKMYQIKSCYKTP 61
 Qy 60 KILSQCH---IRLYAAVNTDCEVFLNGKVGHNIEVLPPE--VDVTKVKGSGENELR 113
 Db 62 LHISEKRYKRVLIYHEGVAAVAQVVLNGLYIGEHKGYTPFDIRIDEVYDWKELNMLT 121
 Qy 114 VVVENLKKVGGPFSKVPDSGHTVGFSGFPFANPFPFGYGIIRVVLBFPDHAILDI 173
 Db 122 VVVDSTER---SDLPKG-----GQIDLYLVGGIYREVSGIYDVVFIKVI 164
 Qy 174 VVDT--SESEPEKLT-----GKVKVIEVSEAVGQEMTIKLGEEBKJRTSN 219
 Db 165 KYETHIYINNEKSLNLIHLENLNNHQSNGVKKVKINDGN-QKEVYK-----ERN 214
 Qy 220 RPEVG-----EFLFN--ARFWSLEDVLYPLKVELE---KDEYTLIDIGRTISWDEK 267
 Db 215 TYLDAVKDYVSFNIENLKDILMDVDNPMIYEIKVGMKINNFSDEYDNKGFREAVFKPD 274
 Qy 268 RLYLNGKPVFLKGFGEHFEFVLGGTFYPLMKIDENILK--WINNSFRTSHYPSSEML 326
 Db 275 GFLNKRKILKGLNRHSGYFVYGAMPRRVEKDAEILKNEHLNLYTSHYPOSKHPL 334
 Qy 327 DLADRLGILVIDEAPVIGITRYHNPETOKIAEDNIRMDHKNHPSVIMSV-ANEP 385
 Db 335 NKCDELGLLVFEIIPG---WQYIGNSEWKKVAEQMLREMITRDMNHPSTILGWVRINSEQ 391
 Qy 386 SNHPDAGEFKALYETANEMDRTPVWVSMMDADERTRDVALKCFDVCNRYGYI 445
 Db 392 ---DDDAFYKMMNKIAHEIDPTROTGVRYI---TNSFLE--DV-----YT 430
 Qy 446 YQGRLEEG--QALEKDIIEELVARRKRPVTEFGADALAGIHYDPMPMSEEV-QAEIVERTIRL 503
 Db 431 FNDLFDHGDINKRLRQGVYTGLEHNVPIYVITYNG-----HMVPTKRDNREKQHEHL 484
 Qy 504 KTIPL--LKKDYIIGHVAFADFKTPQNVRRPILNHK-----GVTRDRROP 549
 Db 485 RHLRIQNASYLDSDISGAIIGWCMFDYNT-----HKDFGSGDRICYHGVMDFRPF 534
 Qy 550 KLVAVYLR 557
 Db 535 KFSYVYK 542

RESULT 11
 D86872
 beta-galactosidase (EC 3.2.1.23) [imported] - *Lactococcus lactis* subsp. *lactis* (strain II
 C:Species: *Lactococcus lactis* subsp. *lactis*
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

A:Reference number: A90981; MUID:84028567; PMID:6313347
A:Accession: A90981
A:Molecule type: DNA
A:Residues: 2-1024 <KAL>
A:Cross-references: UNIPROT:P00722; GB:V00296; NID:G41901; PIDN:CAA23573.1; PID:G1197203
A:Note: translation of initiator Met is not shown
J.Fowler, A.V.; Zabin, I.
J Biol. Chem. 253: 5521-5525, 1978
A:Title: Amino acid sequence of beta-galactosidase. XI. Peptide ordering procedures and
A:Reference number: A92233; MUID:78218239; PMID:97298
A:Accession: A92233
A:Molecule type: protein
A:Residues: 2-1024 <FOM>
A:Note: this is the final paper in a series
R.Calor, M.P.; Miller, J.H.
Nature 285, 38-41, 1980
A:Title: Molecular consequences of deletion formation mediated by the transposon Tn9.
A:Reference number: A93224; MUID:80188189; PMID:6246435
A:Accession: A93224
A:Molecule type: DNA
A:Residues: 356-476 <CAL>
A:Ruteshouser, E.C.; Richardson, J.P.
J. Mol. Biol. 208, 23-43, 1989
A:Title: Identification and characterization of transcription termination sites in the E
A:Reference number: S06878; MUID:89362462; PMID:2475637
A:Accession: S06878
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-147 <RUT>
A:Cross-references: EMBL:X16313; NID:G41903; PIDN:CAA34380.1; PID:G41904
R.Mitryukov, N.N.; Petrov, N.A.; Karginov, V.A.; Vassilenko, S.K.
Bioorg. Khim. 6, 1735-1736, 1980
A:Title: Nucleotide sequence of a lambda-plac 5-1 DNA region coding for a COOH-terminal
A:Reference number: I41218
A:Accession: I41218
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 967-971, 'R', 973-1022, 'E', 1024 <MTK>
A:Cross-references: GB:M83832; NID:G146061; PIDN:AAA23835.1; PID:G146062
R.Jacobson, R.H.; Zhang, X.; Dubose, R.F.; Matthews, B.W.
Submitted to the Brookhaven Protein Data Bank, July 1994
A:Reference number: A65162; PDB:1BGL
A:Contents: annotation; X-ray crystallography, 2.50 angstroms, residues 4-1024
R.Jacobson, R.H.; Zhang, X.J.; Dubose, R.F.; Matthews, B.W.
Nature 369, 761-766, 1994
A:Title: Three-dimensional structure of beta-galactosidase from E. coli.
A:Reference number: A58594; MUID:94277211; PMID:8008071
A:Contents: annotation; X-ray crystallography, 2.50 angstroms
R.Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H64761
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1024 <BIAT>
A:Cross-references: GB:AE000141; GB:U00096; NID:G1786532; PIDN:AACT3447.1; PID:G1786539
A:Experimental source: strain K-12, substrain MG1655
R.Prentki, P.
Gene 122, 231-232, 1992
A:Title: Nucleotide sequence of the classical lacZ deletion delta M5.
A:Reference number: I40987; MUID:93083990; PMID:1339377
A:Accession: I40987
A:Molecule type: DNA
A:Residues: 1-11,43-50 <RES>
A:Cross-references: EMBL:X58252; NID:G40882; PIDN:CAA41206.1; PID:G40883
C:Genetics:
A:Gene: lacZ
A:Map position: 8 min
C:Complex: homotetramer
C:Function:
A:Description: catalyzes hydrolysis of lactose into galactose and glucose

C:Superfamily: beta-galactosidase
C:Keywords: glycosidase; homotetramer; hydrolase; magnesium
F:2-1024/Product: beta-galactosidase #status experimental <A>
F:51-218/Region: alpha complementation
F:51-218/Domain: 1, jelly-roll type-barrel #status predicted <DM1>
F:219-334/Domain: 2, fibronectin type-III fold #status predicted <DM2>
F:335-627/Domain: 3, distorted TIM barrel #status predicted <DM3>
F:628-737/Domain: 4, fibronectin type-III fold #status predicted <DM4>
F:738-1024/Domain: 5, anti-parallel beta-sandwich #status predicted <DM5>
F:417,419,462/Binding site: magnesium (Glu, His, Glu) #status experimental
F:462,504,538/Active site: Glu, Tyr, Glu #status predicted
Query Match 13.8%; Score 414; DB 1; Length 1024;
Best Local Similarity 28.4%; Pred. No. 3.7e-20;
Matches 132; Conservative 59; Mismatches 172; Indels 102; Gaps 16;
QY 15 LNWVNLVETYSKRPPIAVPGSVNE-----QYODLCYE--EG 48
DB 55 LNWENR--AMFAPAEVPESSWLECDLPADTVVSNMGMGYDAPITVNTVPTVNP 112
QY 49 PFT-----YTTTYVPRK-LSQKHRLYPAAVNTDCEVFLNGEKVGNHLEYLPFV 99
DB 113 PFVPTENPTGCSLTFTNVDSWLEQGTIRIIFGVNSAFHLMCGHWVGQDSRLPSEF 172
QY 100 DVTGKVKSGENELRVVENRLKVGSPKVPDGTHTVGFSGFPFANPFFPYGIIIP 159
DB 173 DLSAFLRAGENRLAVWV-LRWSDSY--LEDO-----DMMMSGIFRD 212
QY 160 VLIETDHAIIIDIVDTSEBEKKLGKVKVIEVSEAVQ-----EMTIKIGEEKK 214
DB 213 VSLIHKPTTQISDFHATRENDFS-----RAVLBAEVQCGELRYLRTVSLVWGQETQ 267
QY 215 IIRSNRPFVEGEF-----LENAFSLDEBYLYPLKVEKEBYTL---- 255
DB 268 VASGTAPFGEELIDEGGYADRVTLRLNVENPCLMSAEIPNLRAVVEHTADGTLIEAE 327
QY 256 --DIGRTISWDEKRLYLNGKPVFLKFGFHEEPVLGGOTFPLMIKPFNLKWINANS 313
DB 328 ACVGVREVIENGILLNGKPELLINGVNRHEHPLHGQVMEQTNVODILMKONNENA 387
QY 314 FRTSHYPYSEWLDLDRIGLIVDEA---PH--VGITYHVNPEYOKIAEDNIRIMDR 368
DB 388 VRCSHPNHPMLWTLCDRGLVYVDANIEHTGMVPMNLTDPRMLPMSEKVTIMVQR 447
QY 369 HKHPSVIMSVANEP--ESNHPDAEGFKALYETANEDRRTPV 411
DB 448 DRNHPSVIITWISLGNESGHGANN-----DALYRWIKSVDPSPV 485
RESULT 14
E90678
beta-D-galactosidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099;
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90678
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Gaaswara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90678
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 11024 <HAY>
A:Cross-references: UNIPROT:Q8X665; GB:BA000007; PIDN:BAB33820.1; PID:G13359854; GSPDB:GT
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECS0397
C:Superfamily: beta-galactosidase
Query Match 13.5%; Score 405.5; DB 2; Length 1024;
Best Local Similarity 25.2%; Pred. No. 1.4e-19;
Matches 159; Conservative 88; Mismatches 216; Indels 169; Gaps 26;

Page Blank (uspto)